UNIVERSITY OF
LINCOLN

Life-like Image Processing for Small Target Motion Detection in Cluttered Dynamic Environments

Hongxin Wang

Doctor of Philosophy

2019

School of Computer Science

University of Lincoln

Life-like Image Processing for Small Target Motion Detection in Cluttered Dynamic Environments

Hongxin Wang

School of Computer Science

University of Lincoln

A thesis submitted in partial fulfilment of the requirements of the University of Lincoln for the degree of Doctor of Philosophy

Supervisor

Professor Shigang Yue

April 2019

Abstract

Discriminating targets moving against a cluttered background is a huge challenge for future robotic vision systems, let alone detecting a target as small as one or a few pixels. As a source of inspiration, insects are quite apt at searching for mates and tracking prey – which always appear as small dim speckles in the visual field. The exquisite sensitivity of insects for small target motion, as revealed recently, is coming from a class of specific neurons called small target motion detectors (STMDs). Some of the STMDs have also demonstrated direction selectivity which means these STMDs respond strongly only to their preferred motion direction. Build a quantitative STMD model is the first step for not only further understanding of the biological visual system, but also providing robust and economic solutions of small target detection for an artificial visual system. This research aims to explore STMD-based image processing methods for small target motion detection against cluttered dynamic backgrounds. The major contributions are summarized as follows.

Three STMD-based neural models are proposed in this research named as directionally selective STMD(DSTMD), STMD Plus and Feedback STMD, respectively. The DSTMD systematically models and studies direction selectivity of the STMD neurons, meanwhile provides with unified and rigorous mathematical description. Specifically, in the DSTMD, a new correlation mechanism is introduced for direction selectivity via correlating signals relayed from two pixels. Then, a lateral inhibition mechanism is implemented on the spatial field for size selectivity of the STMD neurons. Finally, a population vector algorithm is used to encode motion direction of small targets. Extensive experiments showed that the proposed DSTMD not only is in accord with current biological findings, i.e. showing directional preferences, but also works reliably in detecting small targets against cluttered backgrounds.

The STMD Plus is developed to discriminate small targets from small-target-like background features (named as fake features) by integrating motion information with directional contrast. More precisely, the STMD Plus is composed of four subsystems – ommatidia, motion pathway, contrast pathway and mushroom body. Compared to existing STMD-based models, the additional contrast pathway extracts directional contrast from luminance signals to eliminate false positive background motion. The directional contrast and the extracted motion information by the motion pathway are integrated in the mushroom body for small target discrimination. The experimental results demonstrated the significant and consistent improvements of the proposed visual system model over existing STMD-based models against fake features.

The Feedback STMD is also designed to filter out fake features by introducing a new feedback mechanism. Specifically, the model output is first temporally delayed then applied to the previous neural layer to construct a feedback loop. By subtracting the feedback signal from the inputs of the STMDs, the background fake features are largely suppressed. Experimental results show that the developed feedback neural model achieves better performance than the existing STMD-based models in discriminating small targets from complex backgrounds.

Declaration

I, Hongxin Wang, declare that this thesis describes an original research carried out on my own. It has not been previously submitted to any university for the award of any degree. Where I have quoted from the work of others, the source is always given.

Acknowledgements

First and foremost, I would like to thank my academic advisor, Professor Shigang Yue, for his tolerance and patience in letting me explore my interest fields. He not only encouraged me to think deeply and creatively, but also taught me how to effectively communicate my research in papers and presentations. I hope through the years I have been able to pick up a little of his ability to see and explain ideas and concepts with such clarity. He will always be a role model as I begin a career in academics.

Professor Jigen Peng was my co-advisor, and I am grateful to him for his help and support when I was in Xi'an Jiaotong University. I especially value his trust in giving me plenty of chances without expecting anything in return. I have learned many things from him: high quality research, effective communication skills and hard-working attitude. I can not ask for more and the high standards he has imparted to me will benefit my research efforts my whole life long.

I would like to take this chance to thank my family for their selfless love and constant support, especially my parents who gave me great support on deciding to start my PhD study and provided me generous help during my study.

The environment at University of Lincoln is full of brilliant and enthusiastic colleagues who have provided me valuable help and discussion. I wish to thank the past and present members in the Computational Intelligent Lab (CIL) who have brought me great convenience on daily life and academic research.

Finally, I appreciate the financial support from School of Computer Science at University of Lincoln and EU FP7 projects EYE2E (269118), LIVCODE (295151), HAZCEPT (318907) and EU Horizon 2020 project STEP2DYNA (691154).

List of Main Publications

[1] **H. Wang**, J. Peng, and S. Yue, "A Directionally Selective Small Target Motion Detecting Visual Neural Network in Cluttered Backgrounds," *IEEE Trans. Cybern.* (***T-CYB***), in press, doi: 10.1109/TCYB.2018.2869384.

[2] **H. Wang**, J. Peng, X. Zheng, and S. Yue, "A Robust Visual System for Small Target Motion Detection Against Cluttered Moving Backgrounds," *IEEE Trans. Neural Netw. Learn. Syst.* (***T-NNLS***), in press, doi: 10.1109/TNNLS.2019.2910418.

[3] **H. Wang**, J. Peng, and S. Yue, "A Feedback Neural Network for Small Target Motion Detection in Cluttered Backgrounds," in *Proc. Int. Conf. Artif. Neural Netw.* (***ICANN***), Oct. 2018, pp. 728-737.

[4] **H. Wang**, J. Peng, Q. Fu, H. Wang and S. Yue, "Visual Cue Integration for Small Target Motion Detection in Natural Cluttered Backgrounds," *Proc. Int. Jt. Conf. Neural. Netw.* (***IJCNN***), 2019, accepted.

[5] **H. Wang**, J. Peng, and S. Yue, "An Improved LPTC Neural Model for Background Motion Direction Estimation," in *Proc. IEEE Int. Conf. Develop. Learn. Epigenetic Robot.*, Sep. 2017, pp. 47-52.

[6] **H. Wang**, J. Peng, and S. Yue, "Bio-inspired Small Target Motion Detector with a New Lateral Inhibition Mechanism," in *Proc. Int. Jt. Conf. Neural. Netw.* (***IJCNN***), Jul. 2016, pp. 4751-4758.

Contents

List of Figures

List of Tables

List of Abbreviations

AMC	Amacrine cell
DSTMD	Directionally selective small target motion detector
EMD	Elementary motion detector
ESTMD	Elementary small target motion detector
FD cell	Figure detection cell
FSTMD	Feedback small target motion detector
HRC	Hassenstein-Reichardt correlator
LGMD	Lobula giant movement detector
LMC	Lobula monopolar cell
LPTC	Lobula plate tangential cell
STMD	Small target motion detector

Chapter 1

Introduction

1.1 Motivation

The dynamic visual world is often complex, with many motion cues at different speeds, directions, distances and orientations, exhibiting various physical characteristics such as size, colour, texture and shape. Being able to detect target motion in the distance and early would put an entity (a robot or an animal) in a good position to prepare for interaction/competition, for example, a flying insect searching for mates in the distance. In the visual world, detecting visual motion in the distance and early often means dealing with small targets with only one or a few pixels in size let alone other physical characteristics (as shown in Fig. 1.1).

Small target motion detection[1] has a wide variety of applications in defences, surveillance, security and road safety. For instance, timely finding micro drones flying over runways and shooting them down could protect airports from disruption. However, detecting small targets against cluttered moving backgrounds is always a challenge for artificial visual systems. The difficulty is reflected in two aspects: first, when a target is far away from the observer/camera, it always appears as a small dim speckle whose size may vary from one pixel to a few pixels in the field of view. In this size, most of physical characteristics, such as color, shape and texture, are difficult to recognize and cannot be used for motion detection. Second, small targets are often buried in

[1] Small target motion detection aims to detect objects of interest which move against cluttered natural environments and appear as small dim speckles. The sizes of small dim speckles vary from 1 pixel to 10×10 pixels, whereas other physical characteristics, such as color, shape and texture, are difficult to recognize and cannot be used for motion detection.

Figure 1.1: An unmanned aerial vehicle (UAV) flying in the distance [1]. Since the UAV is far away from the camera, it appears as a small dim speckle.

cluttered backgrounds and difficult to separate from small-target-like background features. In addition, ego-motion may bring in further difficulties to small target motion detection.

How to detect small target motion in cluttered moving backgrounds robustly with limited resources? Research in insects' visual system have revealed one effective solution. Detecting small targets in naturally cluttered backgrounds is critical for many insect species to search for mates or track prey. As the result of millions of years of evolution, the small target motion detection visual systems in insects are both efficient and reliable [10], [24], [25]. Dragonflies, as an example, can pursue small flying insects with successful capture rates as high as 97% relying on their well evolved vision system [26], [27]. Compared to the visual systems of primate animals, insects' visual systems achieve amazing capability using relatively simple structures and a small number of neurons. Insects' visual pathways are ideal models for designing artificial vision systems for small target motion detection.

In the insect's visual system, a class of specific neurons, called small target motion detectors (STMDs), have been identified as showing exquisite selectivity for small targets (size selectivity) [17], [28], [29]. These STMD neurons give peak responses to targets subtending $1^\circ - 3^\circ$ of the visual region, with no response to larger bars (typically $> 10^\circ$) or to background movements represented by wide-field grating stimuli [28]. In addition, some STMD neurons are directionally selective (direction selectivity) [13],

[30]. They respond strongly to small target motion oriented along a preferred direction, but show weak or no, even fully opponent response to null-direction motion. Null direction is 180° from the preferred direction. Although the postsynaptic pathways of the STMD neurons are still under investigation [31]–[33], it clear that knowing the small target motion and its motion direction at the same time is an advantage in tasks such as tracking mates or intercepting prey.

The electrophysiological knowledge about the STMD neurons revealed in the past few decades, makes it possible to propose quantitative models. Wiederman *et al.* [34] proposed elementary small target motion detector (ESTMD) to account for size selectivity of the STMD neurons. However, the ESTMD did not consider direction selectivity and showed no different responses to small target motion oriented along different directions. To address this issue, Wiederman and O'Carroll [35] mentioned that two hybrid models, i.e., elementary motion detector (EMD)-ESTMD and ESTMD-EMD, could exhibit both size and direction selectivities. In the follow-up studies [36]–[38], these two hybrid models are used for target tracking. Although size selectivity was investigated and direction selectivity was noted in these STMD models, there are still four aspects of limitations.

1) The existing STMD models have not provided unified and rigorous mathematical description.

2) Wiederman *et al.* [34], [35] and Bagheri *et al.* [36]–[38] focused on size selectivity, tracking mechanisms and non-directionally selective properties e.g. velocity and contrast tuning. Since direction selectivity has not been systematically studied, characteristics and performance of the directionally selective STMD models, are unclear.

3) The existing STMD models have not shown the capacity for encoding motion direction of small targets.

4) The existing STMD models cannot discriminate small moving targets from small-target-like background features, which means that their detection results may contain a large number of false positives.

To overcome the limitations of the existing STMD models, this thesis develops three STMD-based neural models with unified and rigorous mathematical description, which are named as Directionally Selective STMD (DSTMD), STMD Plus and Feedback STMD, respectively. Specifically, the DSTMD is devised to model the specific STMD neurons with direction selectivity, and its directionally selective and nondirectionally selective properties are systematically studied and tested. The directionally selective outputs of the DSTMD are used to encode motion direction by a population vector algorithm. To eliminate false positive background motion, the STMD Plus and Feedback STMD are developed by visual cue integration mechanism and feedback loop, respectively.

1.2 Organization of the Thesis

This thesis consists of six chapters. Chapter 1 gives the motivation and organization of the dissertation. Chapter 2 summarizes the related work on modelling motion-sensitive visual neurons. The main contributions of this thesis are contained in Chapter 3, 4 and 5, which presents the designed DTSMD, STMD Plus, and Feedback STMD, respectively. In each of these three chapters, we will discuss the design motivation, describe the proposed model, and demonstrate the experimental results. Chapter 6 finally summarizes the main contributions of this thesis and outlooks the possible future work. The details in each chapter are summarized as follows.

Chapter 2 mainly reviews the related work on motion-sensitive neural modelling. It first introduces the biological findings on three widely investigated motion-sensitive neurons which show exquisite sensitivity to looming, wide-field motion, and small target motion, respectively. Then the computational models of the three neurons and their applications are separately summarized. In addition, we briefly discuss traditional motion detection and small target motion detection approaches in engineering.

Chapter 3 presents a directionally selective STMD-based neural model (DSTMD) for detecting small moving targets and estimating motion directions. It first introduces the background and the overview of this study, and then formulate the newly developed neural model. Finally, the experimental results are presented and discussed. This

chapter is extended based on the publication [1] on the page V.

Chapter 4 proposes a visual system model (STMD Plus) for discriminating small moving targets from background fake features. It first introduces the background and indicates the problem. Following that, the structure and formulation of the proposed visual system model are detailed. Finally, the experimental results are demonstrated and discussed. This chapter is an enriched version of the contents published in [2] on the page V.

Chapter 5 develops a feedback STMD-based neural model (Feedback STMD) for small target discrimination against cluttered moving backgrounds. It first introduces the background and indicate the shortages of the existing STMD-based models. Then the developed feedback loop is described and formulated in details. After that, we present the experimental results and discussions. This chapter is extended based on the content published in [3] on the page V.

Chapter 6 summarizes the main research contributions of this dissertation and discusses the potential work that can be further done in the future.

Chapter 2

Background

In the natural world, object motion always carries significant biological significances. For example, fast approaching predators always appear as rapidly expanding visual stimuli in the visual filed which can trigger the escape of prey [39], [40]. Motion perception is particularly important for the survival of many animal species in critical moments, such as to detect predators or to hunt for prey. As the result of millions of years of evolution, the animals' visual systems are efficient and robust to perceive various motions in cluttered dynamic environments. The exquisite sensitivities of animals for different motions, as revealed in the biological research, are coming from a large number of specialized visual neurons, including small target motion detectors (STMDs) [13], [18], lobula plate tangential cells (LPTCs) [41]–[43], and lobula giant movement detectors (LGMDs) [44]–[46].

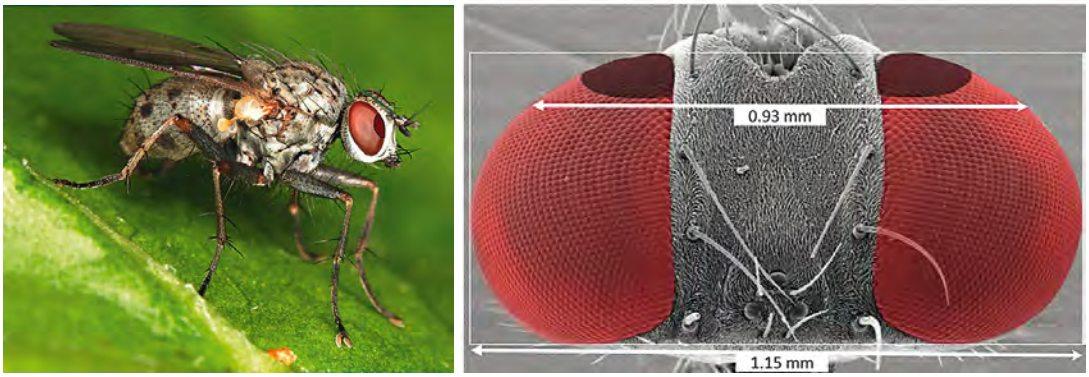
This chapter mainly reviews the related work on three widely investigated motion-sensitive neurons which show preference to looming stimuli [1], wide-field motion [2], small target motion [3], respectively. It begins with a brief introduction about the organization of the animals' visual systems in Section 2.1. Following that, the biological findings on the three motion-sensitive neurons and their computational models are separately summarized and discussed in Section 2.2, 2.3, and 2.4. Finally, Section 2.5 and 2.6 briefly summarize traditional motion detection and small target motion detection approaches in engineering, and discuss the inability of the traditional detection meth-

[1] Looming stimuli stand for the expansion of objects in the visual field.
[2] Wide-field motion denotes the motion of objects or background which occupy large parts of the visual field.
[3] Small target motion refers to the motion of objects which appear as small speckles in the visual field.

ods for small target detection against cluttered natural environments.

2.1 The Insects' Visual Systems

In contrast to vertebrates, insects have compound eyes which consists, depending on the species, of many hundreds to thousands of individual ommatidia (see Fig. 2.1). Visual information perceived by the ommatidia is further processed in several layers of neuropil including lamina, medulla, lobula and lobula plate, as illustrated in Fig. 2.2(a). These neural layers collectively form the optic lobe, where the lamina and the medulla are arranged sequentially, while the lobulla and lobula plate are linked to the medulla in parallel, together forming the lobula complex. Fig. 2.2(b) presents the internal structures of the retina, lamina and medulla. Similar to the retina, the lamina and medulla each are built from about 750 repetitive, retinotopically arranged units, called cartridges and columns, respectively. The neural layers contain roughly 100 different cell types, most of which exist once per unit (ommatidium/cartridge/column). The major cells types of each layer are separately introduced in Section 2.1.1, 2.1.2, 2.1.3 and 2.1.4.



(a) (b)

Figure 2.1: The compound eyes of the fly [2].

2.1.1 Retina

The compound eyes (or called the retina) of the insects are composed of a number of ommatidia. As can be seen in Fig. 2.3, each ommatidium has its own little lens and houses a set of photoreceptors which are grouped into six outer photoreceptors,

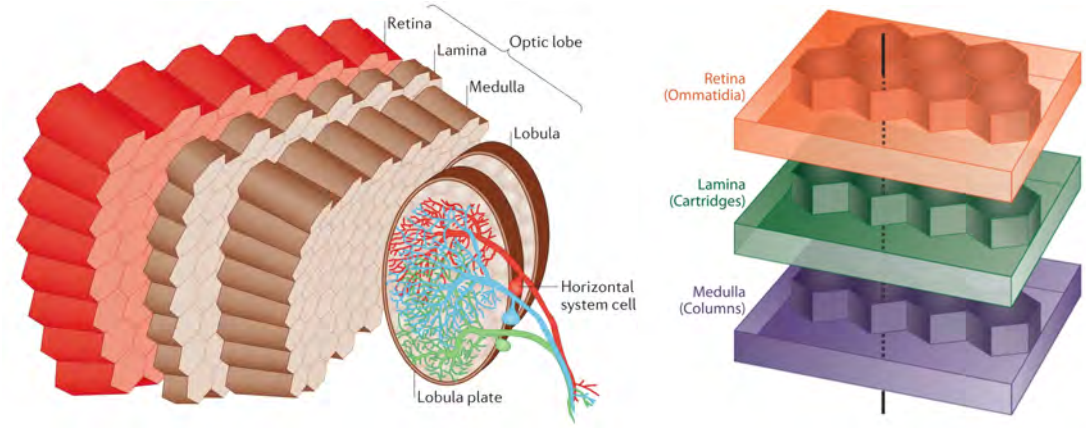


Figure 2.2: (a) The fly optic lobe receives visual input from the ommatidia of the compound eye (in red) and consists of several retinotopically arranged neuropils, called lamina, medulla, lobula and lobula plate [3]. (b) In the fly visual system, spatial relations among successive stages are strictly determined. Retinal ommatidia overlie laminar cartridges, which in turn overlie medullay columns [4].

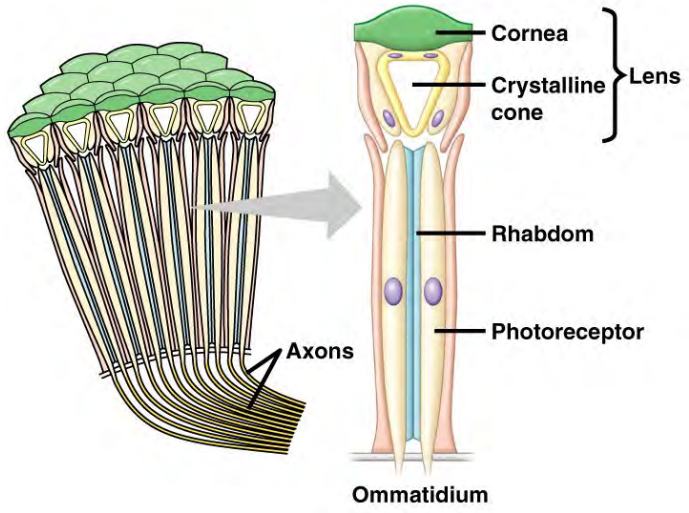
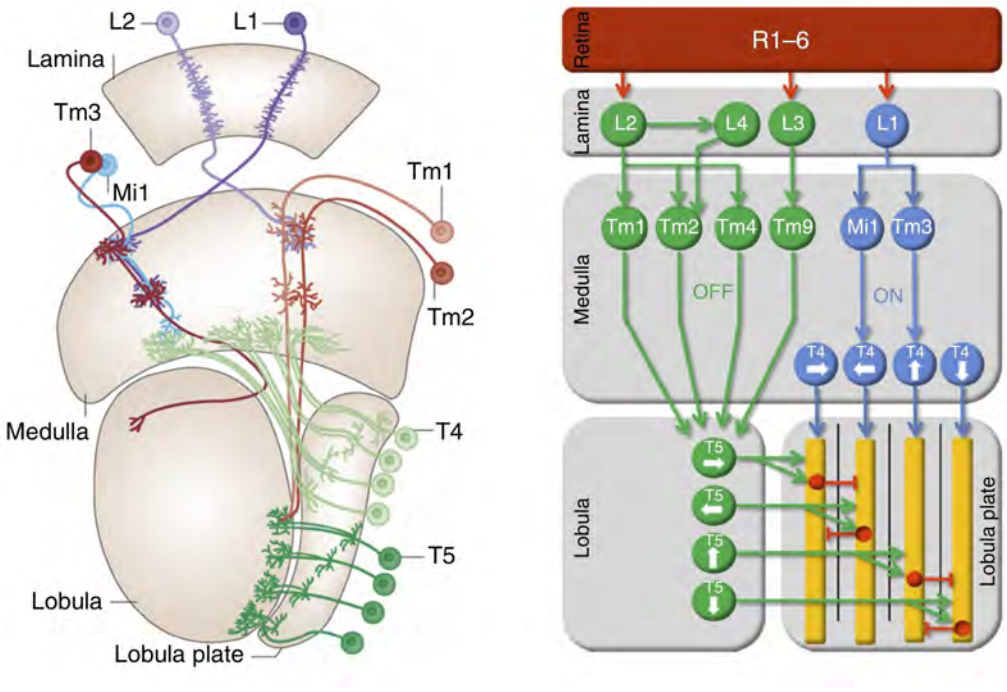


Figure 2.3: The basic anatomy of a single ommatidial unit [5].

R1-R6, and two inner, R7 and R8 [3], [4], [6]. The photoreceptors in each ommatidium receive photons from only a narrow angle, giving rise to a single image point. Therefore the spatial resolution of the insect eyes is largely determined by the angular separation between neighboring ommatidia [6]. Comparing to the vertebrate eyes, the spatial resolution of the insect eyes is much lower, but they can provide insects with a panoramic view of the world.

Inner and outer photoreceptors all act as visual information receptors, however, they differ with respect to spectral sensitivity: R1-R6 show broad spectral sensitivity and support achromatic vision such as motion vision [47], [48], whereas inner pho-



(a) (b)

Figure 2.4: (a) Schematic morphology of main cell types forming the motion pathway in the fly optic lobe [6]. (b) Circuit diagram of the motion pathway of the fly optic lobe [6].

toreceptors (R7 and R8) express more narrow spectral sensitivities, supporting color vision [49]. In addition, the R1-R6 neurons relay visual information to the successive lamina layer, while the R7 and R8 neurons directly provide signal to the medulla layer [50].

2.1.2 Lamina

The lamina contains five different monopolar cell types L1-L5 [51], which are illustrated in Fig. 2.4. As can be seen, each of the R1-R6 is linked to L1-L3 neurons in a single cartridge. Each cartridge, in turn, receives inputs from one each of the R1-R6 photoreceptor cell types [4].

Object motion always induces the spatial-temporal luminance changes in the retina. The large monoplar cells (LMCs) have been demonstrated to be sensitive to these motion-induced luminance changes. More precisely, the L1 and L2 hyperpolarize in response to luminance increments and depolarize in response to decrements [52]–[55]. That is, the L1 and L2 may serve as temporal high-pass filters (or band-pass filters) to

extract information about both luminance increase and decrease from the visual signals provided by the photoreceptors [56], [57].

The L1 and L2 function similarly and their responses show no difference. However, the outputs of the L1 and L2 are applied to two parallel information processing channels, one for luminance decrements and the other for luminance decrements [58]–[60]. This point has been revealed in the biological experiment: when the synaptic output from the L1 is blocked genetically, tangential cells of the lobula plate fail to respond specifically to moving brightness increments while their responses to moving brightness decrements are unaffected; Conversely, if the synaptic output from the L2 is blocked, the tangential cells no longer respond to moving brightness decrements while their responses to moving brightness increments are intact [53].

2.1.3 Medulla

The medulla layer contains more than 60 different cells which can be roughly grouped into medulla intrinsic (Mi) neurons and trans-medulla (Tm) neurons [3]. Fig. 2.4 displays the identified medulla neurons involved in motion detection. As it is shown, Tm1, Tm2 and Tm4 neurons receive inputs from the L2 cells while Tm3 and Mi1 are postsynaptic to the L1 cells.

In medulla layer, two parallel channels selectively respond to luminance increase and decrease caused by object motion [53], [54]. Further research indicates that these two pathways are implemented by Tm1, Tm2, Tm3 and Mi1 [7]. Specifically, the Mi1 and Tm3 neurons respond selectively to brightness increments, with the response of the Mi1 delayed relative to the Tm3; In contrast, the Tm1 and Tm2 neurons respond selectively to brightness decrements, with the response of the Tm1 delayed compared with the Tm2.

The Tm1, Tm2, Tm3 and Mi1 neurons are not directionally selective [7], [61], [62]. Fig. 2.5 presents the responses of the four neurons to light and dark bars moving in different directions. As it is shown, all four neurons exhibit a strong response to moving bars and the amplitude of the responses is independent of motion direction. This demonstrates that directional selectivity of the insects' visual systems may arise

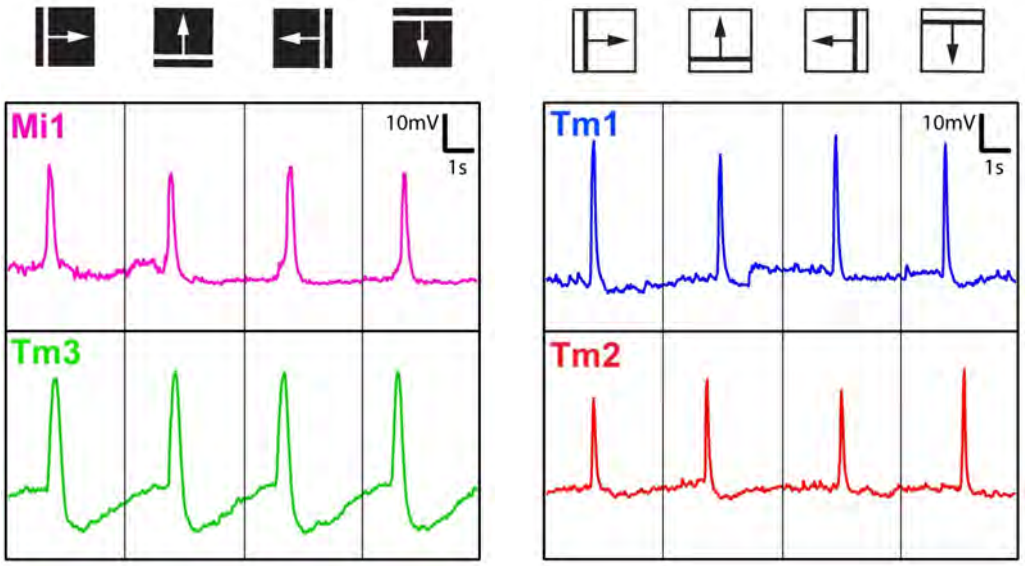


Figure 2.5: Mi1, Tm3, Tm1 and Tm2 are not directionally selective [7]. (a) Top: response of an Mi1 neuron to a white bar moving rightwards, upwards, leftwards and downwards at $100°/s$ on a dark background. Bottom: same as top for a Tm3 neuron. (b) Top: response of a Tm1 neuron to a black bar moving rightwards, upwards, leftwards and downwards at $100°/s$ on a light background. Bottom: same as top for a Tm2 neuron.

in deeper neural layer, i.e., the lobula.

2.1.4 Lobula

The connection from the medulla to the lobula layer are provide by T4 and T5 cells [63]–[65], as we can see from Fig. 2.4(a). To be more precise, Mi1 and Tm3 synapses on the dendrites of the T4 cells whereas Tm1 and Tm2 synapse onto the dendrites of T5 cells. That is, T4 cell receives input from the Mi1 and Tm3, then provide its output to the lobula, while T5 cell collects the outputs of Tm1 and Tm2. T4 cells are found to be selective to luminance increase while T5 cells respond exclusively to luminance decrease [41]. In addition, both T4 and T5 cells are directionally selective, each of which contains four distinct types responding selectively to one of four cardinal directions, i.e., front-to-back, back-to-front, upward, downward (see Fig. 2.4(b)).

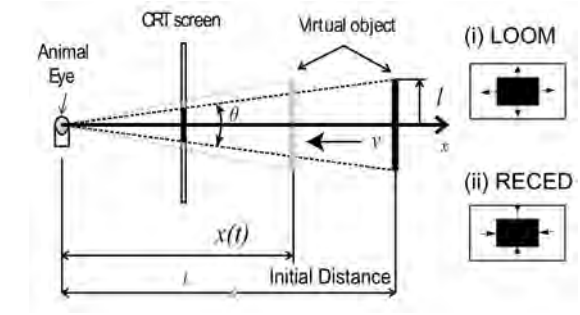
The lobula layer contains numerous specialized cells which integrate signals from previous layers and respond to different features of the visual world. The best characterized motion-sensitive neurons include lobula plate tangential cells, small target

motion detectors, locust giant motion detectors and rotation-sensitive neurons which selectively respond to wide-field motion, small target motion, and looming, respectively. These three neural types are introduced in the following sections.

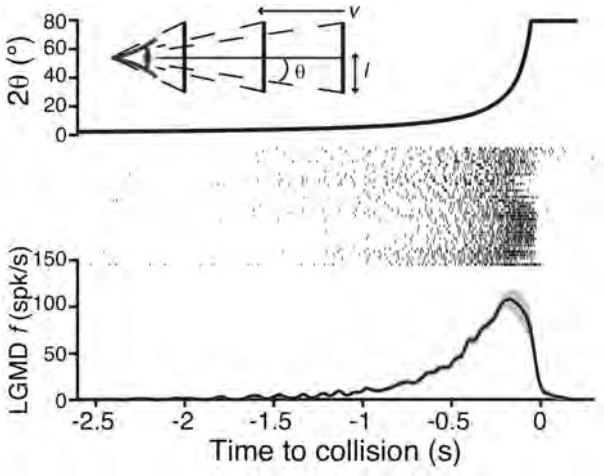
2.2 Looming-sensitive Neurons

Biological Background: A group of large interneurons called the lobula giant movement detectors (LGMDs) has been found in the the lobula layer of locusts [44]–[46]. These LGMDs respond most strongly to looming stimuli over other kinds of visual challenges like receding and translating movements [66]–[68]. Fig. 2.6(a) presents the experimental setup to record LGMD neural responses to visual stimuli. As it is shown, looming stimuli always induce expansion of visual patterns in the retina, whereas receding stimuli result in contraction of patterns. In the biological experiments, expanding and contracting visual patterns displayed on the CRT screen are used to simulate looming and receding objects to elicit the LGMD responses. The responses of the LGMDs to the looming stimuli are illustrated in Fig. 2.6(b). It can be seen that the LGMDs fire a vigorous burst of action potentials to the looming object, where the strength of the response is positively correlated to the angular subtense of the object (2θ).

The LGMD neurons have similar neuromorphology as well as neural characteristics and functionalities, but they still show small differences. For example, LGMD1 [44] and LGMD2 [45], two well characterized LGMD neurons where the LGMD2 neuron is a neighbouring partner to the LGMD1, differ in their looming selectivity. More precisely, the LGMD2 neuron is only sensitive to darker looming objects whilst not responding to oncoming brighter objects against dark background. Compared with the LGMD2 neuron, the LGMD1 neuron can respond to either dark or light approaching objects [45]. The recent research [68] which has investigated the pre-synaptic neuropile layer of the medulla in the locusts' visual systems, found that the specific looming selectivity of the LGMDs is formed well in the pre-synaptic fields. Despite of this, the underlying mechanisms of the LGMDs remain largely unknown.



(a)



(b)

Figure 2.6: (a) Looming and receding corresponded to optical flow of expansion and contraction [8]. (i) Looming stimuli: l was the half-size of the object, v was the approach speed, L was the initial distance and θ is the angular size of the object in degrees. (ii) Receding stimulus. (b) LGMD responses to looming stimuli [9]. Top, schematic of visual stimulus. Note the non-linear increase in angular subtense (2θ), characteristic of looming stimuli. Middle, spike rasters of the LGMD responses to looming stimuli. Bottom, mean instantaneous firing rate (f) of LGMD looming response. Shaded area is $\pm$ SEM.

Computational Model: Rind and Bramwell [67] proposed a functional neural model based on the LGMDs input circuitry. This neural model showed the same selectivity as the LGMD neuron for approaching rather than receding objects and responded best to objects approaching on collision rather than near-miss trajectories. The expanding edges of colliding objects and the lateral inhibition were the key features computed by the model. However, this LGMD-based neural model was either challenged only by pure computer generated visual stimuli [67], or tested in a simple, structured environment [69]. In addition, the success rate of the neural model for collision avoidance was still unsatisfying, about 69% in the tested occasions [70]. To increase the success rate in dealing with colliding objects against a complex background, Yue and Rind [10] further developed a LGMD-based collision detecting neural model with a new mecha-

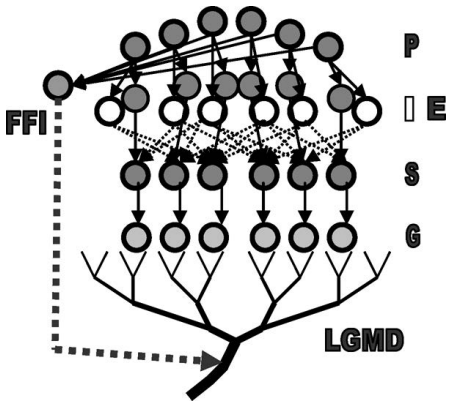


Figure 2.7: Schematic illustration of the LGMD-based neural model for collision detection [10]. There are five groups of cells and two single cells: Photoreceptor cells (P); excitatory and inhibitory cells (E and I); summing cells (S); grouping cells (G); the LGMD cell; and the feed forward inhibition cell (FFI). The input of the P cells is the luminance change. Lateral inhibition is indicated with dotted lines and has one frame delay. Excitation is indicated with black lines and has no delay. The FFI also has one frame delay. The input to FFI is luminance change from photoreceptor cells.

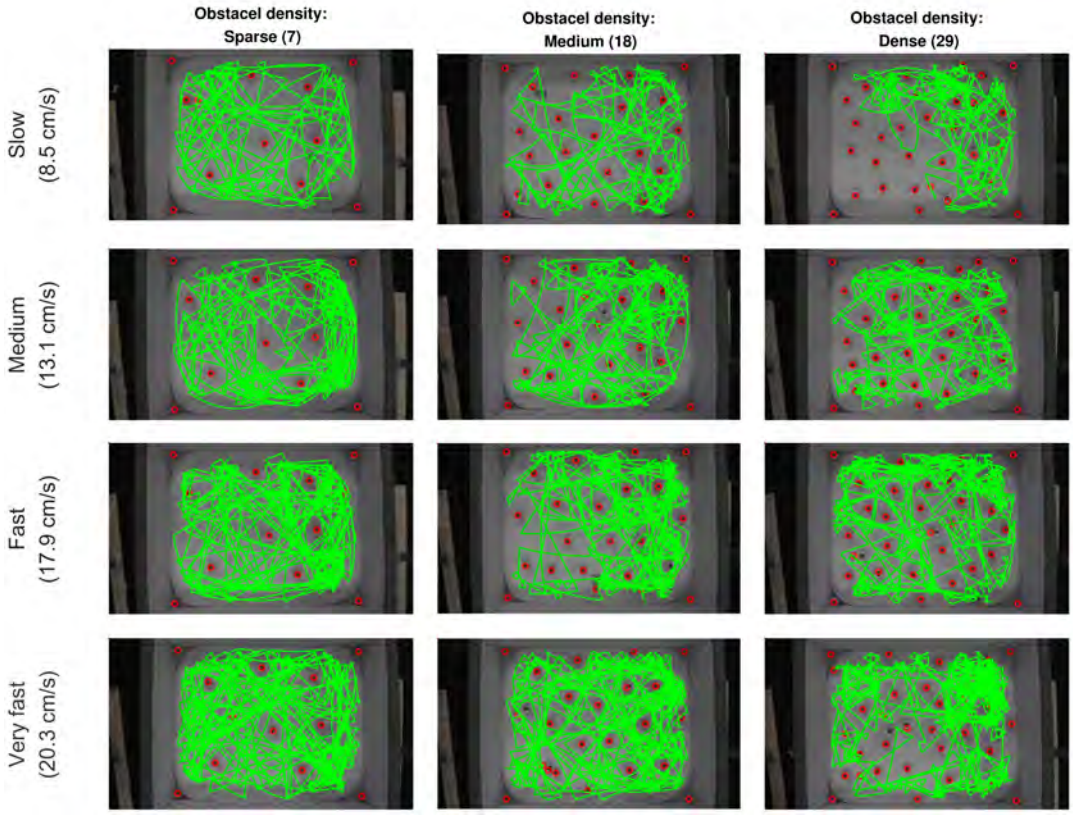


Figure 2.8: Sample of trajectories of the robot in each experiment in the ”forest” [11]. The green lines represent the trajectories of the robot. The initial place of the obstacles are shown as red circles.

nism processing the excitations (refer to pixels with higher value hereafter) before the LGMD cell gathers excitations (see Fig. 2.7). The new mechanism favors grouped excitations by enhancing them and allowing them to reach to the next layer without

decay; however, the isolated excitations is not enhanced and is, therefore, subject to decay. When embedded in a vision-based autonomous micro-robot [11], the neural model has demonstrated its low computational complexity and ability to work under extremely bright or dark conditions (see Fig. 2.8). To enhanced the selectivity for looming and reduce the response to receding and translating objects, Fu *et al.* [71] introduced the ON and OFF pathways for parallel processing excitations and inhibitions. Apart from these models, a number of derivatives based on the work of Rind [67] has also been proposed by introducing new methods to enhance the collision selectivity for approaching objects [72], new layers to reduce environmental noise [73], new mechanism to enhance the performance in high degree of complexity required scenarios [74]. These models have been implemented in hardware like the FPGA [75], as well as applied to collision detection on cars [76] and mobile robots [77].

In contrast, Gabbiani *et al.* [78]–[80] pointed out that the selectivity of the LGMDs for looming stimuli is shaped by the non-linear interactions between the excitations and inhibitions rather the lateral inhibition supported by Rind *et al.* [67]. In addition, the calculations of feedforward excitation/inhibition may be closely related to the angular speed/size of looming objects within visual filed [81]. Based on these points of view, a variant of LGMD models have been proposed to reproduce neural responses and solve the problem of collision detection. Keil [82], [83] developed a mathematical LGMD model to reveal the relationship of the LGMD responses with the angular speed/size and power laws. Badia *et al.* [84] incorporated the non-linear elementary motion detectors in the neural model to sense and avoid potential collision. This model was further improved to fit the non-linear properties of the LGMDs [85], and exhibited invariance of collision detection to looming stimuli with varied shapes, textures and approaching angles [86]. Stafford *et al.* [87] also applied similar strategies to construct collision detecting LGMD model and applied it in driving scenarios.

Summary: The general concept of the above LGMD-based neural models for collision detection is comparing the summation of luminance changes over the whole visual filed with a preset threshold. To calculate luminance changes, these models always

assumes the background is static, which is difficult to satisfy in real applications such as unmanned vehicles. Although these neural models can detect looming objects, they are incapable of discriminating small target motion from cluttered moving background, and do not show size selectivity and direction selectivity.

2.3 Wide-field-motion-sensitive Neurons

Biological Background: In the lobula plate, large neurons have been found integrating locally directionally selective T4 and T5 neural outputs from many hundreds or thousands of medulla columns [88], [89]. These neurons called lobula plate tangential cells (LPTCs), are sensitive to wide-field motion and send motion information to the central brain [41]–[43]. The LPTCs can be generally categorized into vertical and horizontal systems, where the LPTCs of the vertical system are most sensitive to vertical motion while the LPTCs of the horizontal system are tuned to horizontal motion [42], [43], [90].

Computational Models: A wide-field LPTC can be modelled by an array of Hassenstein-Reichardt correlators (HRC), each of which focuses on a small part of the visual field [91]. The HRC model also called (EMD), was inferred from quantitative behavioral studies of the beetle [92]. The schematics of the HRC and fully opponent HRC models are illustrated in Fig. 2.9. As it is shown, the HRC model produces a large response to an object moving along the preferred direction, utilizing the multiplication of two spatially separated signals generated by object motion, one of which has been delayed in time. In contrast, the null-direction motion cannot elicit the response of the HRC model, since the null-direction motion always results in reverse chronological order of the the two input signal, which means that they are unable to coincide at the same time at the multiplication step (see Fig. 2.9(b)). To produce responses to objects moving along both preferred and null directions, two HRC models are combined together in a mirror-symmetric manner as displayed in Fig. 2.9(c)-(d). The output of the fully opponent HRC model is defined by the subtraction of the two mirror-symmetric subunits, which is positive for preferred-direction motion and negative for null-direction

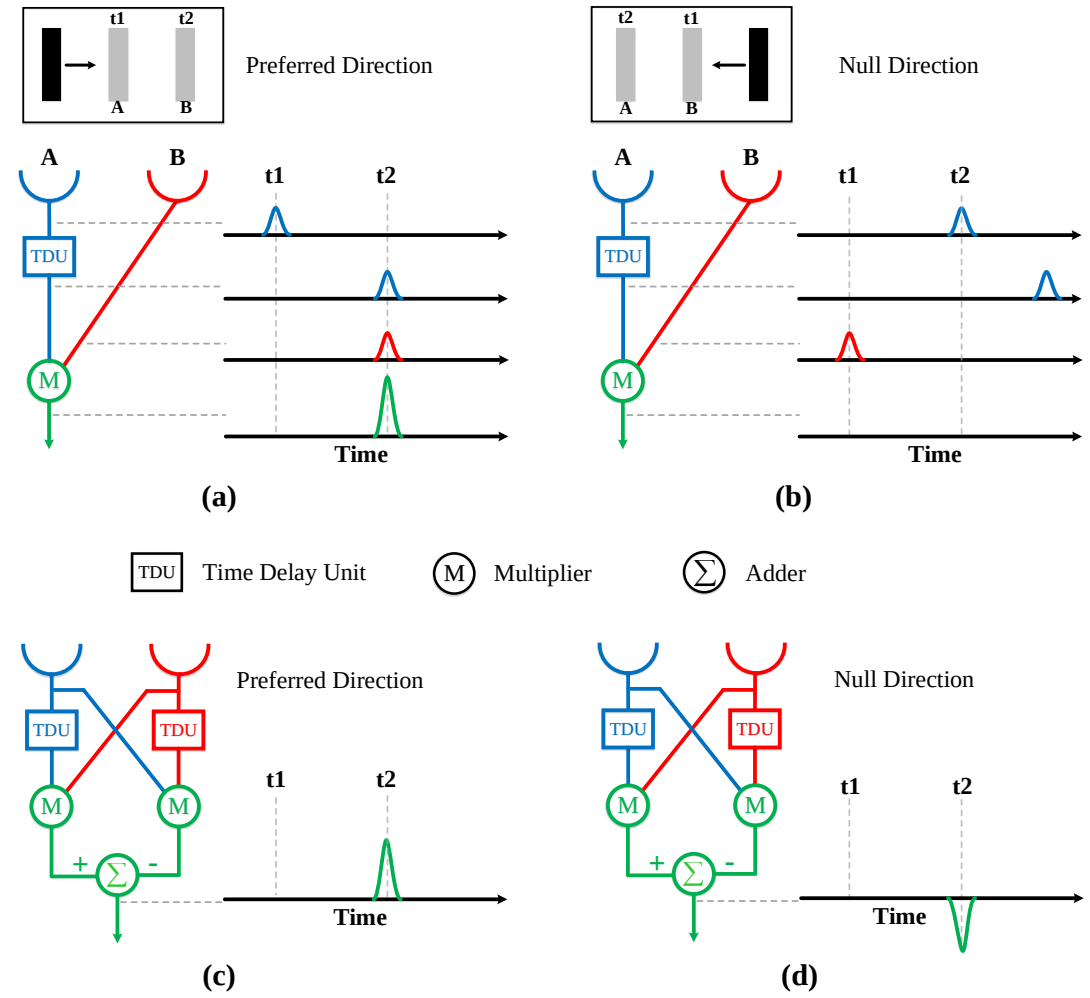


Figure 2.9: Schematics of motion detection of the HRC model and fully opponent HRC. The top panel shows a bar moving in either preferred direction (left panels) or null direction (right panels). The bar crosses the two gray areas (A, B) at two sequential times (t1, t2). (a) For the HRC model, when the bar moving in the preferred direction passes the first photoreceptor, it generates a signal, which is then delayed such that it arrives at a downstream multiplier element coincident with the signal from the bar passing the second photoreceptor; these signals are then multiplied to generate a larger signal. (b) When the bar is moving in the null direction, the two signals of the HRC do not arrive at the same time, and the multiplication step does not generate a response. (c)-(d) Fully opponent HRC model consists of two mirror-symmetrical subunits to generate a positive signal for preferred-direction motion and negative signal for null-direction motion.

motion.

The HRC model is able to faithfully replicate the observed behaviors of the insects [92], but the neural circuits implementing this computation remains under investigation. Recent biological research has identified two parallel channels which separately process luminance increase (ON) and decrease (OFF) signals in the insect's visual systems [7]. To incorporate this new finding, the input signal to the HRC model

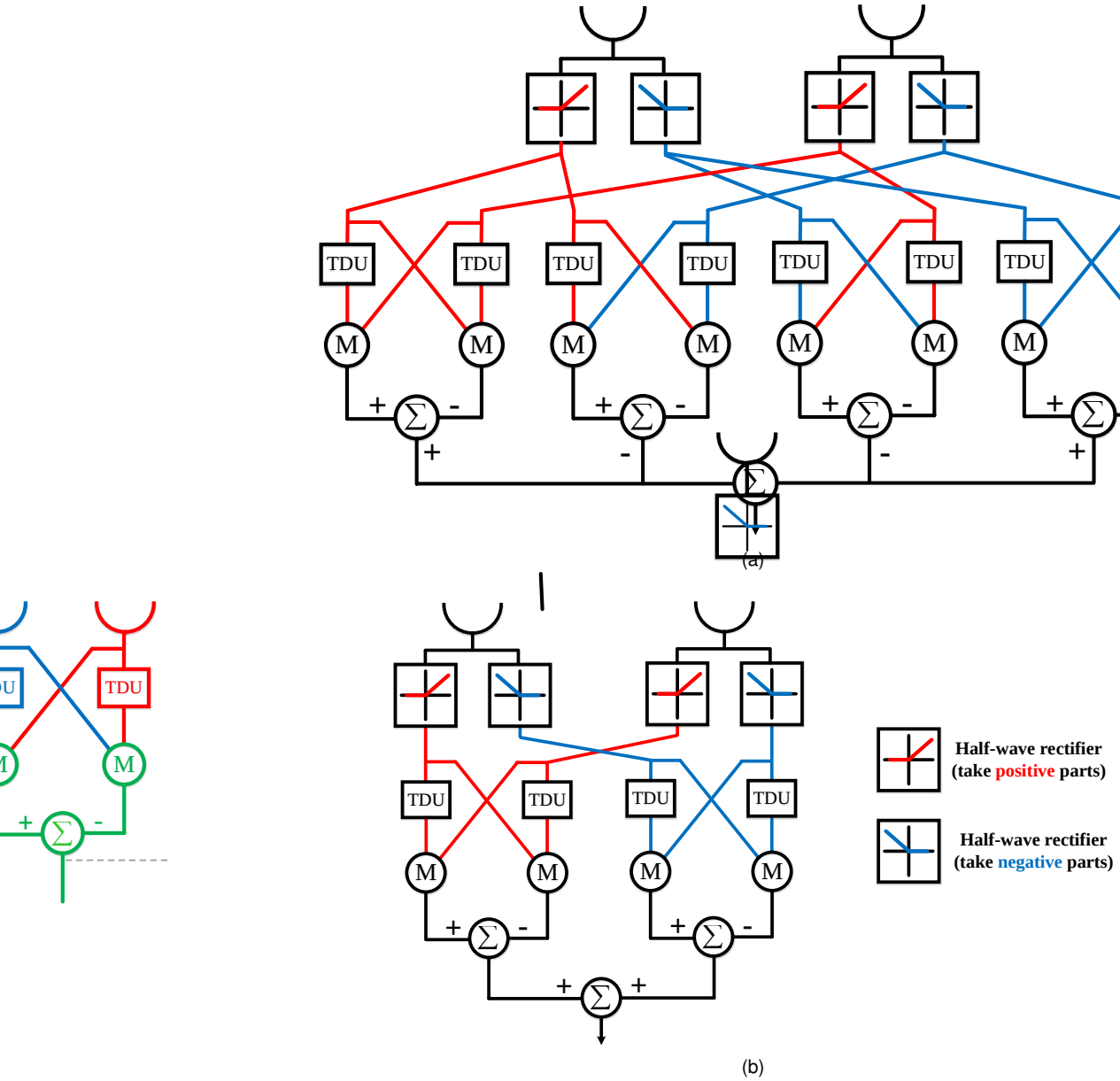


Figure 2.10: (a) A 4-Quadrant-Detector model [12]. Splitting the input into luminance increase (ON) and decrease (OFF) components leads to four parallel channels, one for each combination of input signals (ON-ON, ON-OFF, OFF-ON, OFF-OFF). Each of the four subunits replicates the structure of the HRC model. This model is mathematically identical to the original HRC model. (b) A 2-Quadrant-Detector [12]. Only combinations of the same-type signals are processed (ON-ON, OFF-OFF).

is separated into luminance increase and decrease components by a half-wave rectifier. The separated components are then processed parallelly and combined in possible pairs, giving rise to two alternative models shown in Fig. 2.10. The first model, the "4-Quadrant-Detector" is composed of four parallel HRC models that cover all four possible combinations of luminance increase and decrease signals from two photoreceptors (ON-ON, ON-OFF, OFF-ON, and OFF-OFF). The 4-Quadrant-Detector is

mathematically equivalent to the original HRC model. The second model called 2-Quadrant-Detector [93], only contains two parallel HRC models to process the same-type signals from two photoreceptors (ON-ON, OFF-OFF). Both 4-Quadrant-Detector and 2-Quadrant-Detector are motion-sensitive, but it is still controversial which model exists in the insect's visual systems. In the further experiments which used apparent motion stimuli [4] as visual inputs, the insects' visual systems has been found exhibiting not only positive responses to ON-ON and OFF-OFF sequences, but also negative responses to ON-OFF and OFF-ON sequences [94], [95]. This finding provides a support for the existence of the 4-Quadrant-Detector. However, Eichner *et al.* indicate that the 2-Quadrant-Detector could also respond to ON-OFF and OFF-ON sequences if a small amount of information about the average luminance level (called DC components) is mixed with luminance change signals as model input [12]. That is, the existence of the 2-Quadrant-Detector cannot be ruled out in the insects' visual systems. In contrast, Clark *et al.* advocate for a Weighted-Quadrant-Detector which contains six weighted HRC models to process six pairs of luminance increase and decrease signals (ON-ON, OFF-OFF, OFF-ON, ON-ON, OFF-OFF, ON-OFF). The Weighted-Quadrant-Detector is sensitive an illusory percept "reverse phi", revealing that it may be widely used for motion detection in the visual systems. [54].

Summary: The HRC model and its extensions are generally termed correlation detectors which compute a spatiotemporal correlation between the input intensities at two or multiple separate locations for motion detection. Correlation detectors have the advantage of an intuitive description in the insects' visual systems, and could be applied in velocity estimation [96]–[98], collision avoidance [99], [100] as well as object tracking [101], [102]. However, their responses depend on not only velocity, but also texture and contrast of objects. Thus, they cannot measure the velocity of objects accurately [99]. In addition, these models are sensitive to object motion, nevertheless they are unable to distinguish small moving objects from large ones.

[4] Apparent motion stimuli denotes that the the brightness in two adjacent bars is stepped sequentially from an intermediate level, that is also present in the surround, to either a high (ON-Step) or to a low (OFF-Step) level.

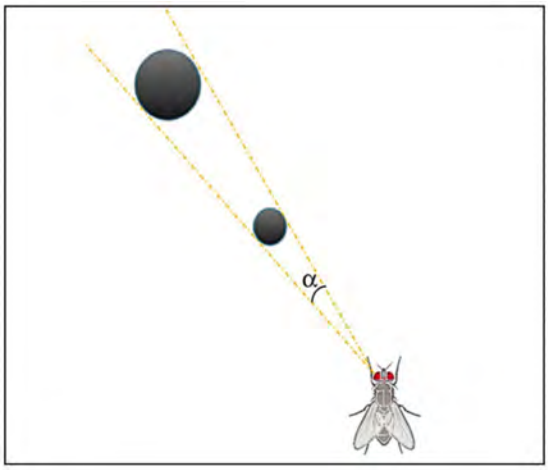


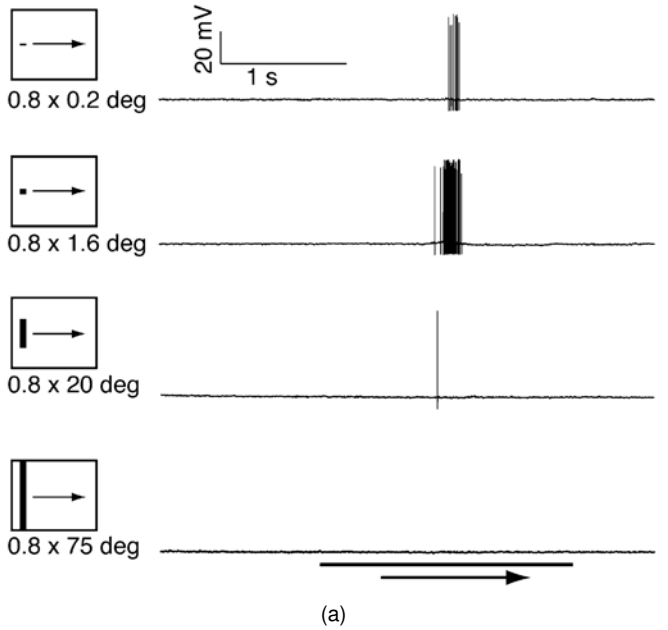
Figure 2.11: An object that is close and small subtends the same size on the retina (α degrees) as one that is far away and big [2].

2.4 Small-target-motion-sensitive Neurons

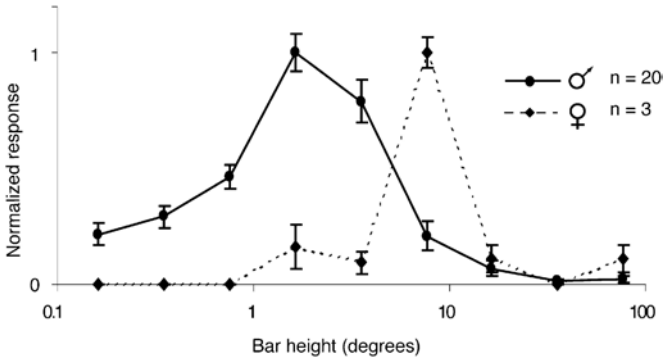
Visual neurons which preferentially respond to small objects have been found in a wide range of animal species, such as cat [103]–[105], monkey [106]–[108], pigeon [109]–[111], toad [112], locust [113], [114], hoverfly [115], hawkmonth [116], [117], dragonfly [30], [118], [119], blowfly [19], [120]. These visual neurons differ in the size of their receptive fields [5] and the preferred size of the objects [21]. For example, the STMD neurons in dragonflies or hoverflies respond most strongly to objects subtending $1^\circ - 3^\circ$ of the visual field (as illustrated in Fig. 2.11) [13], [18]. With the increase of object size, the STMD neural response will gradually decrease, and eventually vanish completely for objects occupying $> 10^\circ$ of the visual field or wide-field stimuli such as grating. As a comparison, other visual neurons like figure detection (FD) cells of blowflies, give peak responses to objects which occupy $6^\circ - 12^\circ$ of the visual field and still may respond, although at a considerably lower level, to wide-field stimuli [121]–[123].

In Section 2.4.1 and 2.4.2, we separately review the related work on two extensively studied small-target-motion-sensitive neurons, i.e., the STMD neurons and FD cells. Each section begins with an introduction about the biological properties of the neuron, then the computational models are summarized and discussed.

[5] The receptive field of an individual visual neuron is the particular region of the the visual field in which a stimulus will modify the firing of that neuron.



(a)



(b)

Figure 2.12: Size selectivity of the STMDs [13]. (a) Raw neural responses of a STMD to bars of four different heights ($0.8°$ wide by $0.2° - 75°$ high) moving through the center of the receptive field at $50°/s$. The soild bars and arrows represent the peri-stimulus duration and direction of target motion. (b) Normalized STMD neural response (mean $\pm$ SEM) to targets of different heights for male (solid line, $n = 20$) and female (dashed line, $n = 3$). n denotes sample number.

2.4.1 Small Target Motion Detectors

Biological Background: Small target motion detectors (STMDs), a class of specialized neurons found in the fly visual system, display exquisite sensitivity to small moving targets against cluttered background [17]. Although the underlying mechanisms for such remarkable small-target sensitivity remain unclear, some basic functional properties of the STMDs have been described, including size selectivity, robustness, direction selectivity and velocity selectivity.

1) Size selectivity implies that the neuron only respond to objects whose sizes

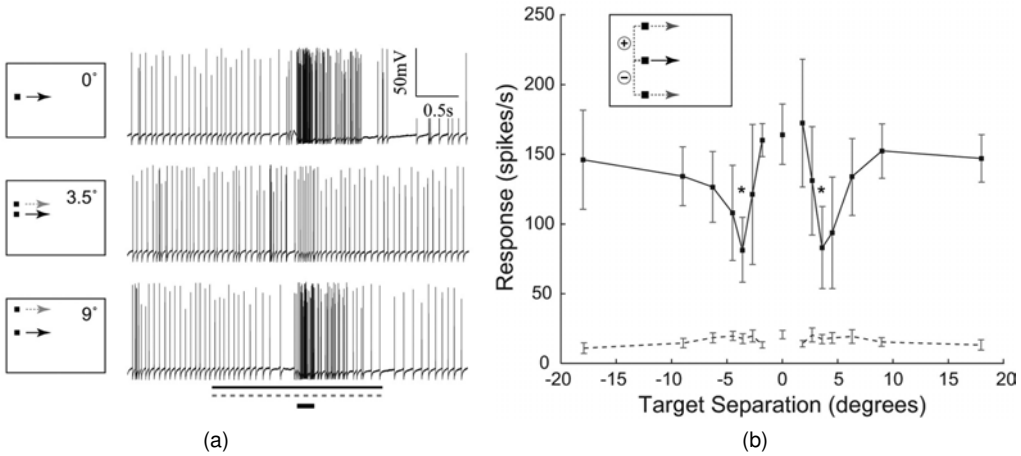


Figure 2.13: Local spatial inhibition [14]. (a) Top: raw spiking response to a single $0.9° \times 0.9°$ target drifting at $55°/s$ from left to right. Middle: the response of the same neuron is attenuated when a distracter target is added at separation (measured center to center). Bottom: the response is restored when the two targets are separated by $9°$. The long bar beneath the trace indicates the peristimulus duration (2 s) of the primary target (solid line) and distracter target (dashed line), and the short bar indicates the 200 ms ($11°$) analysis window. (b) Neural responses with respect to target separation [mean (N) $\pm$ SEM, $N = 8$], where the responses were determined by averaging the spike rates within a 200 ms analysis window. In each trial, the distracter target was separated vertically from the primary target as indicated by the x-axis (distances measured center to center). Negative values indicate that the distracter target drifted below the primary target; positive values indicate that it drifted above (see pictogram). At a separation of $0°$, the distracter target overlaid the primary target. The spontaneous rate is shown by the dashed line.

are in a specific range. To demonstrate size selectivity of the STMDs, black moving bars with different heights are used as visual stimuli in the experiment. As shown in Fig. 2.12(a), the STMD neuron fires a vigorous burst of action potentials only when a optimal-sized target ($0.8° \times 1.6°$) passes through the receptive field. The responses of the STMD to larger targets ($0.8° \times 20°$ and $0.8° \times 75°$) are much weaker or even vanish completely. Fig. 2.12(b) presents the normalized STMD neural responses with respect to the bar height. As can be seen, the STMDs exhibit strong size selectivity, where the male STMD has an optimum target height of $1.6°$, while the female STMD shows a preference for larger target height, about $8°$. The selectivity of the STMDs for small moving targets, as revealed in [14], may come from a lateral inhibition mechanism. To reveal the role of the lateral inhibition in size selectivity, the STMD neural responses to a primary target with a distracter target at varying separations are

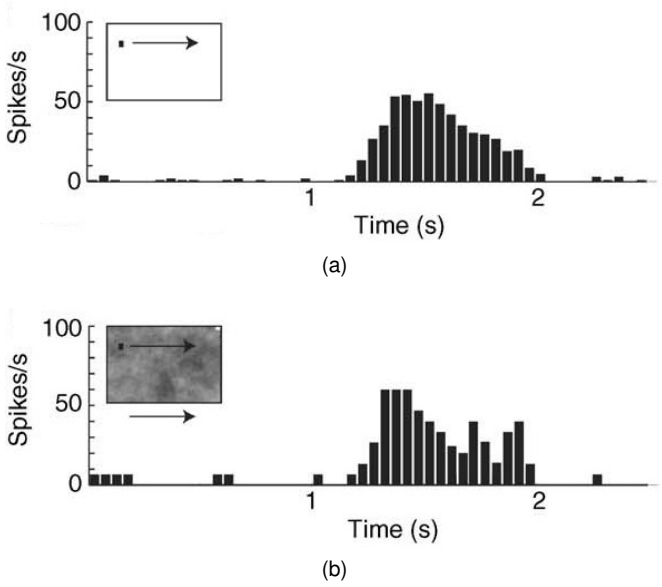


Figure 2.14: The responses of STMDs to a small target against blank background and cluttered background [15]. (a) The response of a STMD to a dark high-contrast target ($0.8° \times 0.8°$) traversing a blank background from left to right at $50°/s$. (b) The response of the same STMD to the same target conditions, but with the blank background exchanged with a cluttered background moving right at the same velocity $50°/s$. The cluttered background has high contrast and spatial statistics similar to those of natural scenes [16].

recorded and presented in Fig. 2.13. As it is shown, the STMD neuron exhibits strong response to the primary target in the case of separation $0°$. However, the response is significantly suppressed by up to 50% when a distracter target is moving at the separation of $3.5°$ to the primary target. As the increase of the separation between the two targets, the neural response gradually restores and reaches the original strength at separation of $9°$. These findings indicate that a lateral inhibition mechanism plays an important role in size selectivity of the STMDs.

2) Robustness means that the STMDs continue to respond robustly to small targets in the presence of background clutter. Fig. 2.14 displays the STMD reponses to a small target in the blank and cluttered backgrounds. Comparing these two subplots, we can find that the STMD responses still remain strong and robust when the small target and the cluttered background move at the same velocity. The relative motion between the object and the background has been considered an important cue for visual neurons to discriminate moving objects, such as fig-

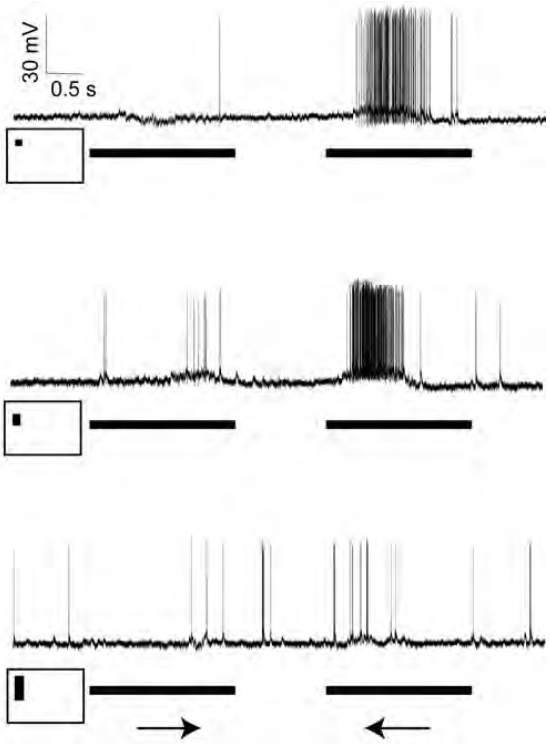


Figure 2.15: Raw neural responses of a STMD to bars of three different heights ($0.8°$ wide by $0.8°, 3°, 15°$ high) moving leftward or rightward [17]. The soild bars and arrows represent the peri-stimulus duration and direction of target motion.

ure detection (FD) cells [6] [120], [125], [126] and object motion sensitive (OMS) cells [127]. These neurons receive inhibition from wide-field motion to achieve selectivity for moving objects, so they can only detect objects which have relative movement to the background [122], [127]–[130]. In contrast, the STMDs are able to discriminate small targets against cluttered background without relative motion, which means that the STMDs do not receive wide-field inhibition and use a different mechanism to discriminate small targets. As revealed in [14], this mechanism may be a local lateral inhibition.

3) Direction selectivity refers to that the neuron only respond to objects that are moving in a preferred direction. Fig. 2.15 presents an example for determining the preferred direction of a STMD neuron. In the experiment, black bars with different heights moving leftward or rightward are used as visual stimuli. As can be seen, the STMD neuron shows a preference for leftward motion. More precisely, the STMD responds vigorously to the leftward motion of the optimal-

[6] In biology, a figure refers to a closed area of visual field wherein the spatiotemporal statistics of motion are distinct from the background [124]. For example, an object moving relative to the background always corresponds to a figure in the visual field.

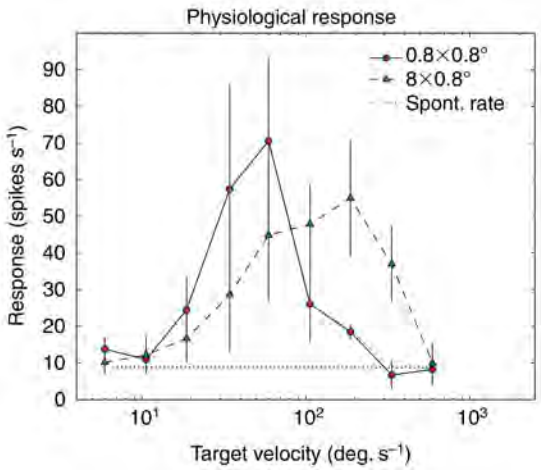


Figure 2.16: The recorded response of the STMDs to target velocity [18]. Two target sizes were used: $0.8° \times 0.8°$ (red circles, solid line) and $8°$ wide $\times$ $0.8°$ high (green triangles, broken line). Error bars denote standard error of the mean (sample number $n = 4$).

size bar ($0.8° \times 3°$), but exhibits almost no response to the bars moving from left to right regardless of their heights.

4) Velocity selectivity stands for that the neuron only respond to objects whose velocities are in a specific range. To demonstrate the velocity selectivity of the STMDs, the neural responses to different target velocities are recorded and illustrated in Fig. 2.16. For the target whose size is equal to $0.8° \times 0.8°$, the STMDs exhibits strong responses to velocities ranging from $6°/s$ to $600°/s$ where the maximum response is reached at $60°/s$. As the increase of target width, the peak has shifted to a higher velocity about $190°/s$.

Computational Models: Wiederman *et.al.* developed elementary small target motion detector (ESTMD) to account for size selectivity of the STMDs. The schematic of the ESTMD is illustrated in Fig. 2.17. In the ESTMD, luminance change signals from a single photoreceptor are separated into increase and decrease components by two half-wave rectifiers. To suppress the luminance changes caused by large objects, the separated luminance increase and decreased are further laterally inhibited. Then the luminance decrease component is temporally delayed and multiplied with the luminance increase component to produce a large response to small dark moving targets. However, the ESTMD is not directionally selective and unable to estimate motion di-

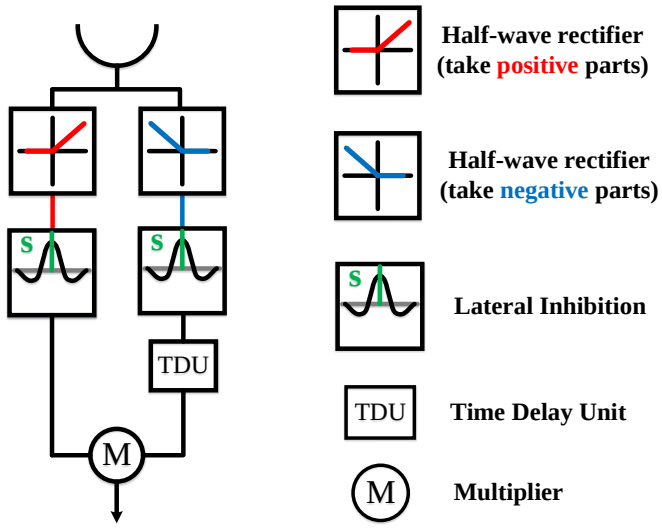
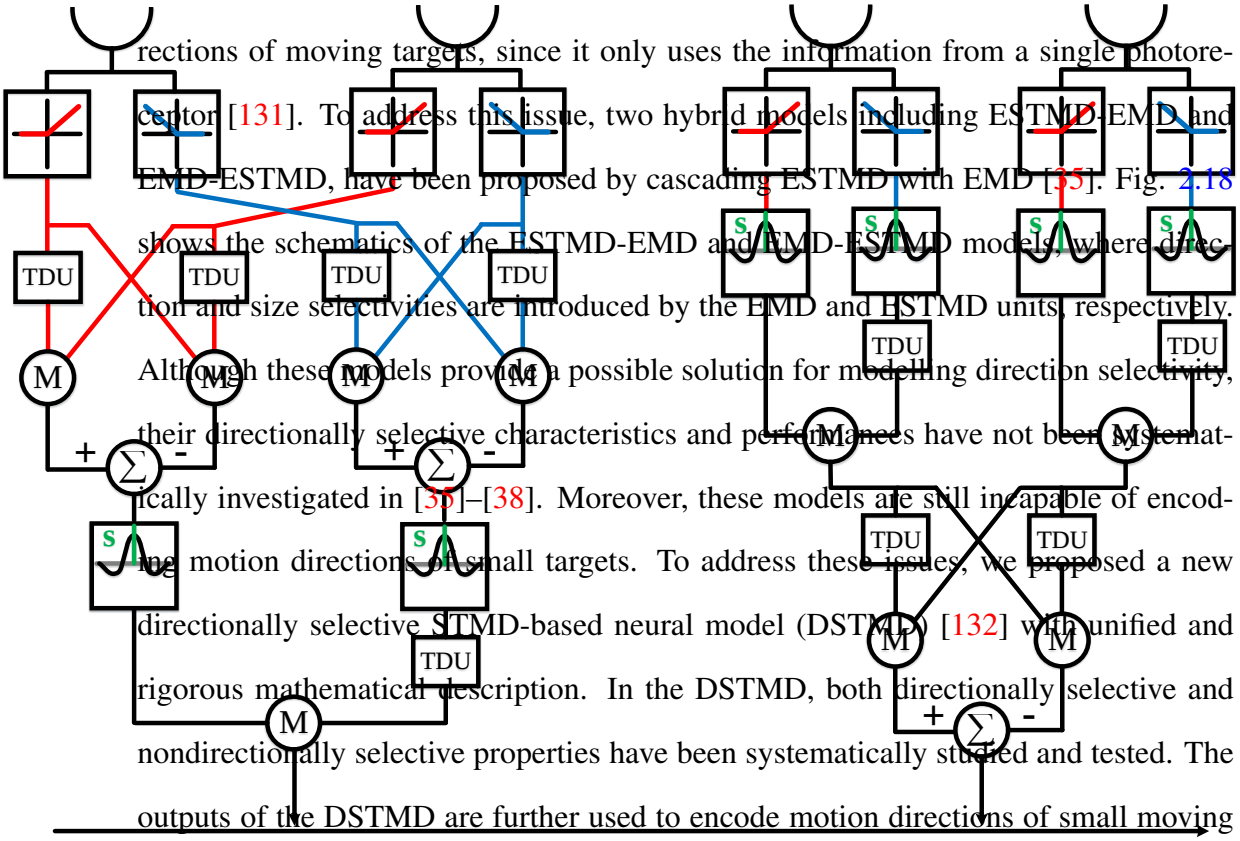


Figure 2.17: Schematic of the ESTMD model. In the model, information about luminance changes from a single photoreceptor is split into luminance increase (positive part) and decrease components (negative part) by two half-wave rectifiers. The two components are further laterally inhibited, finally multiplied together where the decrease component is temporally delayed.



rections of moving targets, since it only uses the information from a single photoreceptor [131]. To address this issue, two hybrid models including ESTMD-EMD and EMD-ESTMD, have been proposed by cascading ESTMD with EMD [35]. Fig. 2.18 shows the schematics of the ESTMD-EMD and EMD-ESTMD models, where direction and size selectivities are introduced by the EMD and ESTMD units, respectively. Although these models provide a possible solution for modelling direction selectivity, their directionally selective characteristics and performances have not been systematically investigated in [35]–[38]. Moreover, these models are still incapable of encoding motion directions of small targets. To address these issues, we proposed a new directionally selective STMD-based neural model (DSTMD) [132] with unified and rigorous mathematical description. In the DSTMD, both directionally selective and nondirectionally selective properties have been systematically studied and tested. The outputs of the DSTMD are further used to encode motion directions of small moving objects by a population vector algorithm. However, the DSTMD and existing STMD-based models only extract motion information for small target detection, so they are unable to discriminate small moving targets from small-target-like background features, meaning that their detection results always contain a number of false positives. To filter out false positives, we developed a visual system model called STMD Plus [133]

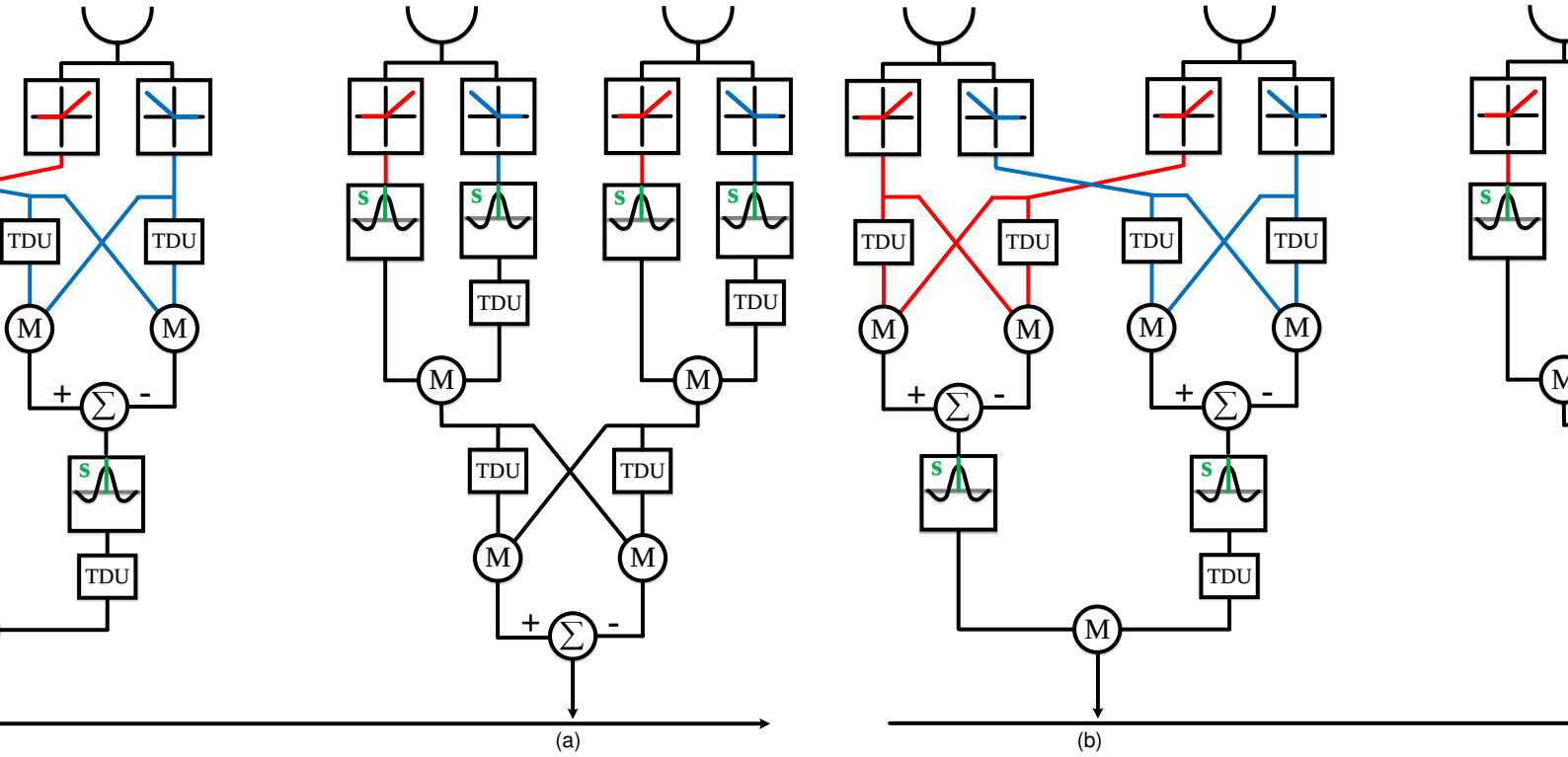


Figure 2.18: Schematics of two hybrid detectors, i.e., ESTMD-EMD and EMD-ESTMD. (a) The ESTMD-EMD model cascades two ESTMDwith an EMD where the outputs of the two ESTMD are fed into the EMD for introducing direction selectivity. (b) The EMD-ESTMD model cascades two EMD with a ESTMD where the directionally selective outputs of the two EMD are applied to the ESTMD for size selectivity.

combining motion information with directional contrast. Specifically, small targets are successfully discriminated from the false positives by comparing the changes of directional contrast on their motion traces. Since the motion trace recording for all detected objects is computationally intensive, we developed a feedback STMD-based model (Feedback STMD) to overcome this limitation [134]. A feedback loop that applies the model output to the previous layer is designed to eliminate background fake features and avoid intensive computational cost for motion trace recording.

Summary: The above models all satisfy the basic properties of the STMD neurons such as size selectivity and/or direction selectivity by similar mechanisms. More precisely, a strong lateral inhibition mechanism formulated as a difference of Gaussians (DoG) is implemented on the model outputs for size selectivity, whereas the signals of luminance changes over time from two different positions are correlated together

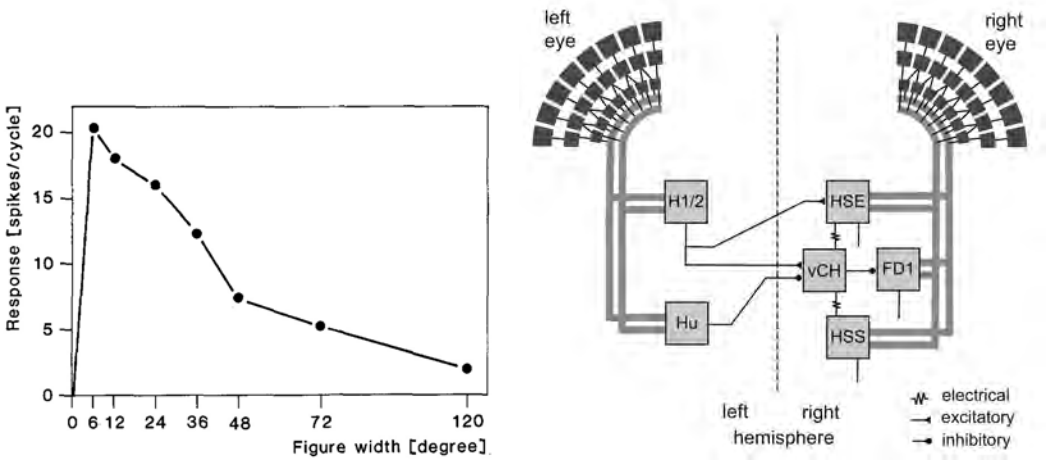


Figure 2.19: (a) Responses of the FD cells with respect to different object widths [19]. (b) Wiring sketch of the FD1 cell input circuit [20]. Motion-sensitive elements of the right FD1 circuit that have a horizontally preferred direction. The FD1 cell and most of its presynaptic elements presumably receive retinotopic motion input (thick gray lines) from large parts of one eye. The right vCH cell inhibits the FD1 cell and receives itself excitatory and inhibitory input from motion sensitive LPTCs of both brain hemispheres. The left H1 and left H2 excite the right vCH cell, whereas the left Hu cell inhibits it. The right HSE cell and the right HSS cell are electrically coupled to the right vCH cell. FD1, HSE, and HSS are output neurons of the optic lobe, whereas H1, H2, Hu, and vCH connect exclusively to other LPTCs.

for direction selectivity. Although these models have demonstrated the reliable ability to detect small moving targets against cluttered backgrounds, it still needs to be noted that they are entirely reliant on an objects contrast against background. They do fail to detect a small moving object if the object has no contrast to its background.

2.4.2 Figure Detection Neurons

Biological Background: Figure detection cells (FD cells), one class of lobula plate tangential cells, have also been demonstrated preferentially responding to small objects [120], [122], [123], [125], [126]. Compared to the STMD neurons, the FD cells prefer larger moving objects. More precisely, the preferred object size of the FD cells ranges from $6°$ to $12°$ [21], [126], as shown in Fig. 2.19(a), whereas the STMD neurons exhibit strongest response to objects as small as $1° - 3°$ [17], [28]. In addition, the FD cells discriminate small moving objects from cluttered background, as long as there is relative motion cues. On the contrary, the STMD neurons can detect small moving objects regardless of whether objects have relative movement to the background.

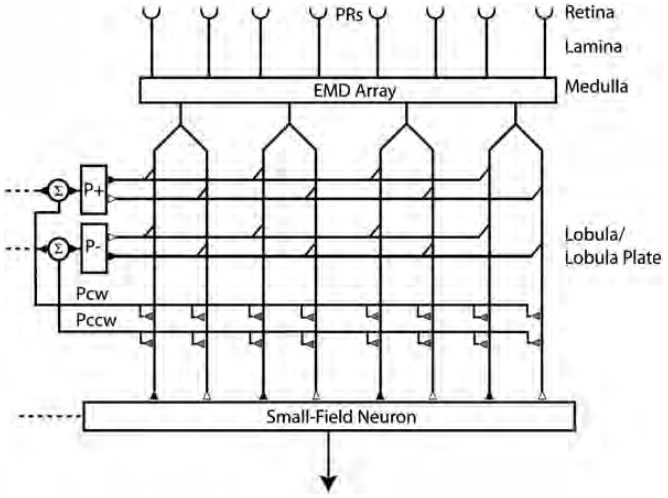


Figure 2.20: Schematic of the small-field model. Luminance signals from photoreceptors (PRs) are input into an array of elementary motion detectors (EMD array). The EMD outputs are split into positive and negative components and summed across the entire visual field by direction-selective monocular pool cells (P+ and P-) and then clockwise (Pcw) and counter-clockwise (Pccw) binocular pool cells. These directionally selective binocular pool cells then interact via shunting inhibition with individual motion-detector output channels, which are then summed by the final small-field-sensitive output unit. Excitatory and inhibitory synapses are shown as black and white triangles, respectively. Shunting inhibition is shown by gray triangles. S indicates a sum. Dashed lines indicate possible contralateral interactions. This figure is adapted from [13].

These results suggest that in contrast with the STMD neurons, the FD cells may use a different mechanism to shape size selectivity. Further research [21], [135] revealed that size selectivity of the FD cells primarily comes from the inhibition from wide-field tangential neurons. Fig. 2.19(b) demonstrates the wiring sketch of the FD cell input circuit. As can be seen, the vCH cell [135] respond to wide-field motion by integrating excitatory and inhibitory inputs from other neurons including HSE, HSS, H1, H2 and Hu cells [136]–[138]. The outputs of the vCH cell are fed into the FD cell to inhibit background motion and mediate responses to small objects. If the inhibition from the vCH cell is removed, the size selectivity of the FD cell will be lost [135].

Computational Models: A few models, called small-field systems [126], [139]–[141], have been proposed to simulate the FD cells. These models are composed of four subsystems including photoreceptors, elementary motion detectors (EMDs), wide-field neuron and small-field neuron, i.e., the FD cell. Luminance signals are received by the photoreceptors, then applied to the array of the EMDs for extracting motion in-

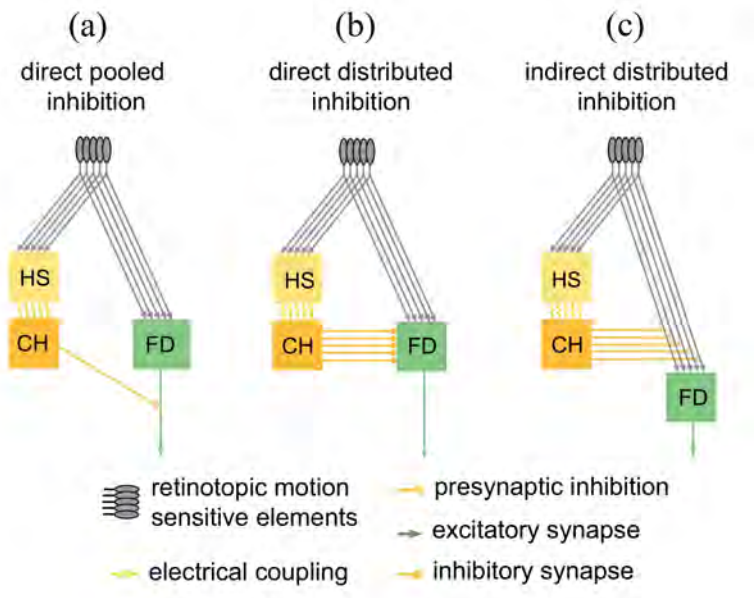


Figure 2.21: Schematics of potential circuits of the input organisation of an FD-cell [21]. The FD-cell receives excitatory retinotopic input from motion sensitive elements. Inhibitory input of the FD-cell is mediated by the vCH-cell via HS-cells. For simplicity, only one of the two HS-cells that provide input to the vCH neuron is shown in this sketch. The coupling between the HS-cells and the vCH-cell is shown to be dendro-dendritic and occurs via gap junctions. (a) The vCH inhibits the FD-cell after spatial pooling (direct pooled inhibition, DPI). (b) The vCH inhibits the FD-cell dendro-dendritically in a distributed way (direct distributed inhibition, DDI). (c) The vCH inhibits the retinotopic input elements of the FD-cell in a distributed way (indirect distributed inhibition, IDI).

formation at each pixel. The wide-field and small-field neurons integrate the outputs of the EMD array to detect background motion and object motion, respectively. The outputs of the wide-field neuron are further fed into the small-field neuron via shunting inhibition. Fig. 2.20 illustrates the schematic of the small-field system. As is shown, the small-field neuron receives input from the EMD array, as well as inhibitory input from the wide-field neuron (Pcw, Pccw) to shape selectivity for small objects. These small-field systems were modified to allow a simulated fly to track a small moving target in a virtual environment [141]. Although these systems exhibit preference for small objects, the spatial distributed interactions in the input circuit of the FD cell remain unclear. To solve this issue, Hennig *et al.* [21] analyzed three wiring schemes of the input circuits of the FD cells, as displayed in Fig. 2.21. The experimental results demonstrate that the latter two wiring schemes, i.e., direct distributed inhibition and indirect distributed inhibition, can account well for size selectivity of the FD cells and the dependence of their responses on the relative velocity between object and background.

To reproduce the time-dependent responses of the FD cell to dynamically naturalistic stimuli, Hennig *et al.* [20] developed a model of the FD and its input circuit, where the model parameters were optimized according to the electrophysiological experiments.

Summary: The above models are able to reproduce properties of the FD cell and detect small moving objects, but their detection performances are highly dependent on velocity difference between moving objects and background. When moving objects and background are relatively static, these FD-based models would be powerless whereas those STMD-based models could still perform well. In addition, the size selectivity of the FD-based models is shaped by the inhibition of background motion. In relatively simple background such as grating, the stable background inhibition can effectively induce size selectivity. However, it would be difficult to be maintained in complex natural backgrounds which cannot provide sustainable and stable inhibition.

2.5 Traditional Motion Detection Methods

Traditional motion detection methods such as optical flow [142], background subtraction [143] and temporal differencing [144], have been developed to detect normal-sized objects like pedestrians and vehicles. They utilize physical characteristics including shape, color and texture, to segment regions corresponding to moving objects from the background. Nonetheless, these methods would be powerless for objects that are as small as one pixel or a few pixels, because it is difficult to identify objects' physical characteristics in such small sizes. Additionally, the above-mentioned methods may not work for cluttered moving backgrounds, as small moving objects could be submerged among the pixel error when applying background motion compensation [145]. Three traditional motion detection methods including optical flow, background subtraction and temporal differencing, are reviewed as follows.

2.5.1 Optical Flow

Optical flow refers to the pattern of apparent motion of objects in visual scene caused by the relative motion between an observer and the objects, and can also be defined as the distribution of apparent velocities of movement of brightness patterns in an image [146]. The definition originates from a physiological description of the visual perception of the world through image formation on the retina [147]. The optical flow estimation methods always assume that the intensity of each pixel remains constant during displacement, i.e., brightness constancy constraint. Specifically, let $I : \Omega \times T \rightarrow \mathbb{R}$ denote an image sequence, where $\Omega \subset \mathbb{R}^2$ and T is the sampled time interval of the sequence. The brightness constancy constraint can be given by

$$\frac{\mathrm{d}I}{\mathrm{d}t}(x, y, t) = 0. \tag{2.1}$$

Using the chain rule for differentiation, we see that,

$$\frac{\partial I}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial I}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial I}{\partial t} = 0. \tag{2.2}$$

If we let

$$u = \frac{\mathrm{d}x}{\mathrm{d}t}, v = \frac{\mathrm{d}y}{\mathrm{d}t} \tag{2.3}$$

then it is easy to see that we have a single linear equation in the two unknowns u and v,

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0. \tag{2.4}$$

The above equation can be rewritten as vector form

$$\nabla I(\boldsymbol{x}) \cdot w(\boldsymbol{x}) + I_t(\boldsymbol{x}) = 0 \tag{2.5}$$

where $w(\boldsymbol{x}) = (u(x), v(x))^\top$, $\boldsymbol{x} = (x, y)^\top$, $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})^\top$, $I_t = \frac{\partial I}{\partial t}$, and $\cdot$ denotes the inner product.

Obviously, two unknown components of optical flow $w(\boldsymbol{x})$ cannot be recovered only by Eq. (2.5), so an additional spatial coherency constraint on $w(\boldsymbol{x})$ needs to be

introduced to make the problem well-posed. To combine with the newly introduced constraint, the brightness constancy constraint is always formulated as a penalty term,

$$\rho_{\text{data}}(\boldsymbol{x}, I_1, I_2, w) = \phi(I_2(\boldsymbol{x} + w(\boldsymbol{x})) - I_1(\boldsymbol{x})) \tag{2.6}$$

where $I_1 = I(\cdot, t)$ and $I_2 = I(\cdot, t+1)$ represent two successive frames; $\phi(\cdot)$ is a penalty function. Since $I_2(\boldsymbol{x} + w(\boldsymbol{x})) - I_1(\boldsymbol{x})$ is the discrete form of $\frac{\mathrm{d}I}{\mathrm{d}t}(\boldsymbol{x}, t)$ which can be approximated by $\nabla I(\boldsymbol{x}) \cdot w(\boldsymbol{x}) + I_t(\boldsymbol{x})$, the above equation is rewritten as

$$\rho_{\text{data}}(\boldsymbol{x}, I_1, I_2, w) = \phi(\nabla I(\boldsymbol{x}) \cdot w(\boldsymbol{x}) + I_t(\boldsymbol{x})). \tag{2.7}$$

The spatial coherency constraint is introduced and imposed in either local or global regions, resulting in two different methods, i.e., parametric approaches [148]–[152] and regularized models [146], [153]–[155].

The parametric approaches restrict the problem in a small subregion of the image $\mathcal{R} \subseteq \Omega$. In the subregion $\mathcal{R}$, the optical flow w is assumed to be characterized by the associated parameter vector $\boldsymbol{\alpha}$. If we restrict ourselves to linear models, then w can be given by

$$w_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \sum_{k=1}^{K} b_k(\boldsymbol{x}) \alpha_k(\boldsymbol{x}) \tag{2.8}$$

where $\boldsymbol{b} = \{b_k(\boldsymbol{x}) | k \in [1, K], b_k(\boldsymbol{x}) \in \mathbb{R}^2\}$ are basis functions and $\boldsymbol{\alpha}(\boldsymbol{x}) = \{\alpha_k(\boldsymbol{x}) | k \in [1, K]\}$ are parameters. Because of the spatial constancy constraint, all pixels in subregion $\mathcal{R}$ have the same motion vector $w(\boldsymbol{x})$, i.e., the same parameter $\boldsymbol{\alpha}$. The problem is transformed to find the optimal parameter $\boldsymbol{\alpha}$ by minimizing the summation ρ_{data} of each pixel in subregion $\mathcal{R}$. That is,

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \sum_{\boldsymbol{x} \in \mathcal{R}} g(x) \rho_{\text{data}}(\boldsymbol{x}, I_1, I_2, w_{\boldsymbol{\alpha}}) \tag{2.9}$$

where $g(x)$ is a weighting function.

The regularized models add a regularization term with the penalty term and mini-

mize their summation on the whole image to find the optimal w. That is,

$$E_{\text{global}}(w) = \int_{\Omega} \rho_{\text{data}}(\boldsymbol{x}, I_1, I_2, w_{\boldsymbol{\alpha}}) + \lambda \rho_{\text{reg}}(\boldsymbol{x}, w) d\boldsymbol{x} \tag{2.10}$$

where λ is a tuning parameter and $\rho_{\text{reg}}(\boldsymbol{x}, w)$ denotes the regularization term. In general, $\rho_{\text{reg}}(\boldsymbol{x}, w)$ is designed to smooth w in regions of coherent motion while preserve discontinuities at the boundaries of moving objects.

2.5.2 Background Subtraction

Background subtraction methods first estimate a reference image without moving objects (called background image) from the temporal sequence of the input frames. Then moving objects is detected by calculating the difference between the current frame and background image. That is,

$$D(x, y, t) = |I(x, y, t) - I_B(x, y, t)| \tag{2.11}$$

where $D(x, y, t)$, $I(x, y, t)$ and $I_B(x, y, t)$ represent the difference image, input frame and background image, respectively. The performance of background subtraction methods is largely dependent on the estimated background image. In dynamic scene, the background image needs to be regularly updated to adapt to varying luminance conditions and geometry settings [156]. A number of methods have been developed to effectively estimate background images. Two of the most popular background modelling techniques are reviewed as follows, including Gaussian Mixture Model (GMM) [157] and Principle Component Pursuit (PCP) [158].

Gaussian Mixture Model: The GMM-based method proposed by Stauffer and Grimson [157] provides a statistical formulation for background modelling where the intensity of each pixel is subject to one (or more) Gaussians. The multi-Gaussian distribution of each pixel is updated for each input frame to accommodate slow lighting variations and objects blending into, or permanently leaving [159]. Specifically, for each pixel, an image generation model M_t is devised from previous measurements $\{Z_0, Z_1, ...Z_{t-1}\}$. The current measurement Z_t is assumed to be generated from M_t,

and the likelihood that this pixel belongs to the background is given by

$$P(\mathbf{Z}_t|\mathbf{M}_t) = \sum_{h=1}^{H} \alpha_h \mathcal{N}(\boldsymbol{\mu}_h, \Sigma_h) \tag{2.12}$$

$$\mathcal{N}(\boldsymbol{\mu}_h, \Sigma_h) = \frac{1}{(2\pi)^{d/2}|\Sigma_h|^{1/2}} e^{-\frac{1}{2}(\mathbf{Z}_t-\boldsymbol{\mu}_h)^\top \Sigma_h^{-1}(\mathbf{Z}_t-\boldsymbol{\mu}_h)} \tag{2.13}$$

where d is the dimensionality of the measurement space which is equal to 3 for RGB image; μ and Σ determine the mean and covariance of each Gaussian; α_h is weighted factor and the their summation equals to 1, i.e., $\sum_h \alpha_h = 1$; $|\cdot|$ denotes the matrix determinant. The channels (RGB or YUV) are considered independent, so the covariance of each Gaussian Σ_h can be written as

$$\Sigma_h = \begin{pmatrix} \sigma^2_{(h,1)} & 0 & 0 \\ 0 & \sigma^2_{(h,2)} & 0 \\ 0 & 0 & \sigma^2_{(h,3)} \end{pmatrix} \tag{2.14}$$

where the subscripts $1, 2, 3$ refer to the channel number.

To determine whether a pixel belongs to the background or foreground, the difference between its current measurement Z_t and the mean μ_h of Gaussian h is calculated and compared with the variance $\boldsymbol{\sigma}_h$ of the Gaussian. That is,

$$||\mathbf{Z}_t - \boldsymbol{\mu}_h|| < k\boldsymbol{\sigma}_h \tag{2.15}$$

where k is set as 2 or 3; $\boldsymbol{\sigma}_h$ is a vector representing the variance Gaussian distribution of index h. The operator is true if all components at the left are smaller than $k\boldsymbol{\sigma}_h$. If Eq. (2.15) is satisfied for a Gaussian h', the pixel is labelled as background and the parameters of Gaussian h' are updated as follows,

$$\alpha'_h \leftarrow (1-\delta)\alpha'_h + \delta \tag{2.16}$$

$$\boldsymbol{\mu}'_h \leftarrow (1-\rho'_n)\boldsymbol{\mu}'_h + \rho'_n \mathbf{Z}_t \tag{2.17}$$

$$\boldsymbol{\sigma}'^2_h \leftarrow (1-\rho'_n)\boldsymbol{\sigma}'^2_h + \rho'_n(\mathbf{Z}_t - \boldsymbol{\mu}'_h)^\top(\mathbf{Z}_t - \boldsymbol{\mu}'_h) \tag{2.18}$$

$$\rho'_n \leftarrow \delta\mathcal{N}(\mathbf{Z}_t|\boldsymbol{\mu}'_h, \boldsymbol{\sigma}'_h) \tag{2.19}$$

where δ determines the adaptation rate. For the other Gaussians $h \neq h'$, the mean and variance remain unchanged, but the α_h is modified as $\alpha_h \leftarrow (1-\delta)\alpha_h$. If Eq. (2.15) fails for all Gaussians, the pixel is labelled as foreground and the parameters of the Gaussian with the small weight α_h is replaced by

$$\alpha_h \leftarrow \delta, \quad \boldsymbol{\mu}_h \leftarrow \mathbf{Z}_t, \quad \boldsymbol{\sigma}_h^2 \leftarrow \bar{\boldsymbol{\sigma}}^2$$

where $\bar{\boldsymbol{\sigma}}^2$ is an initial high variance.

A number of methods have been developed to improve the GMM-based background subtraction. For example, Zivkovic and Van Der Heijden [160] developed a adaptive GMM which can automatically select the needed number of Gaussians for each pixel; Yi *et al.* [161] used the mixture of dual-mode Gaussians to model background, meanwhile prevent the background model from being contaminated by the foreground; Chen *et al.* [162] designed a hierarchical and block-based background subtraction method by combining the GMM with a contrast histogram.

Principle Component Pursuit: In the past decade, principle component pursuit (PCP) methods have been one of the leading techniques for background modelling and widely used to detect moving object in videos captured by static cameras [163]. The most PCP methods are always based on two assumptions: (1) the background parts in an image sequence contain highly redundant information which lies in a low dimensional subspace. That is, the background parts can be approximated by a low-rank matrix $\mathbf{B}$ whose column vector represents the background of each frame; (2) The foreground moving objects deviate from low-rank subspace and can be approximated by a sparse matrix $\mathbf{F}$. To obtain a low-rank matrix and a sparse matrix from the image sequence, the PCP [164] is adopted and the problem can be formulated as follow,

$$\min_{\mathbf{B},\mathbf{F}} ||\mathbf{B}||_* + \lambda||\mathbf{F}||_1 \quad \text{w.r.t.} \;\; \mathbf{X} = \mathbf{B} + \mathbf{F} \tag{2.20}$$

where $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n}] \in \mathbb{R}^{p \times n}$ is an input image sequence of n frames and each $\mathbf{x}_i \in \mathbb{R}^p$ denotes the i-th frame with p pixels; $||\mathbf{B}||_*$ represents the nuclear norm (sum of the singular values of $\mathbf{B}$), $||\mathbf{F}||_1$ stands for the l_1-norm (sum of the absolute values

of all the entries in $\mathbf{F}$), and $\lambda = 1/\sqrt{max(p,n)}$.

Based on the above general concept, various algorithms have been proposed and successfully applied for detecting object motion [165]–[167]. However, these methods cannot be applied to process image sequences captured by moving camera since the matrix $\mathbf{B}$ is no longer low-rank. To address this issue, a transformation matrix τ is applied to the matrix $\mathbf{X}$ for global motion compensation. In order to find $\mathbf{B}$, $\mathbf{F}$ and τ, Eq. (2.20) is revised as follow

$$\min_{\mathbf{B},\mathbf{F},\tau} ||\mathbf{X}^{\circ}\tau - \mathbf{B} - \mathbf{F}||_F + \lambda||\mathbf{F}||_1 \quad \text{w.r.t. } \mathbf{X}^{\circ}\tau = \mathbf{B} + \mathbf{F} \tag{2.21}$$

where $||\cdot||_F$ is the Frobenius norm. Along this line, some methods have been developed to improve the detection performances. For instance, Zhou *et al.* [168] considered the contiguity of moving regions in the formulation; Chen *et al.* [169] used a spatio-temporal coherency of consecutive frames in the optimization formula; Ebadi *et al.* [170] adopted the block sparse structure of the pixels corresponding to the moving objects in their formulation.

2.5.3 Temporal Differencing

Temporal differencing method calculate the pixel-wise differences between two consecutive frames to detect moving object [171], [172]. That is,

$$D(x,y,t) = |I(x,y,t) - I(x,y,t-1)| \tag{2.22}$$

where $I(x,y,t)$ and $I(x,y,t-1)$ denote the input frames t and $t-1$, respectively; $D(x,y,t)$ represents the absolute difference between $I(x,y,t)$ and $I(x,y,t-1)$. Comparing the obtained $D(x,y,t)$ with a given threshold, moving regions can be roughly determined.

Temporal differencing is only able to estimate moving regions in stationary background. To deal with camera or background motion, images should be aligned before detecting moving objects. Generally, corner points in each frame are first detected, then used to match corresponding points in consecutive frames for image alignment.

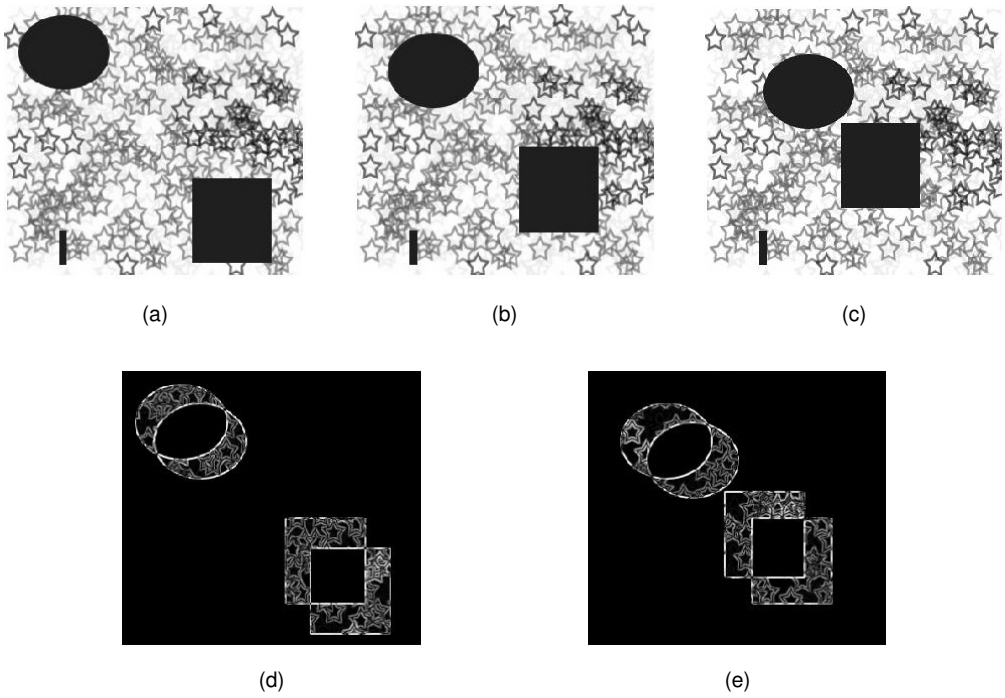


Figure 2.22: Original images: (a) the 1st frame $I(x, y, t-1)$, (b) the 2nd frame $I(x, y, t)$, and (c) the 3rd frame $I(x, y, t+1)$. Results of two-frame differencing: (d) $|I(x, y, t) - I(x, y, t-1)|$, (e) $|I(x, y, t+1) - I(x, y, t)|$. This figure is adapted from [22].

In addition, temporal differencing is adaptive to dynamic environments, but incapable of extracting complete shapes of moving objects [173]. Fig. 2.22 presents the results of two-frame differencing. As can be seen, two motion blobs are generated by each moving object in the difference image, where the two blobs are located at the object's position in the current and previous frames, respectively. Besides, the intersection of the two blobs contains a hole without any luminance changes. The above phenomenon is called ghosting effect [172] and could be avoid by considering the difference of three consecutive frames. The difference image for three consecutive frames is given by the minimum of the difference image between frame t and $t-1$ and the difference image between frame t and $t+1$ [174]. That is,

$$D(x, y) = \min(|I(x, y, t) - I(x, y, t-1)|, |I(x, y, t) - I(x, y, t+1)|). \quad (2.23)$$

As have mentioned above, consecutive frames need to be aligned before calculating their pixel-wise differences. However, the imprecise image registration will generate additional noise. In order to suppress noise, the minimum differences of each pixel in

small neighborhoods was adopted in [175]. The minimum difference between a pixel in the current frame I_t and all pixels (x_i, y_j) in the neighborhood N is calculated by

$$D(x,y) = \min_{i,j \in N}(|I(x,y,t) - I(x_i,y_j,t-1)|, |I(x,y,t) - I(x_i,y_j,t+1)|). \quad (2.24)$$

2.6 Infrared Small Target Detection

Previous research and application of small target detection has mainly focused on infrared images. These infrared-based methods strongly rely on significant temperature differences between the background and objects of interest, such as rockets, jets and missiles. However, such significant temperature difference is rare in natural world. Moreover, the detection environment of these methods were mainly sky and/or ocean, which are much more clear and homogeneous than the cluttered natural environments. These infrared-based methods may not work in a natural environment with lots of bushes, trees, sunlight and shadows, let alone to meet the needs of compact in size and low energy consumption in real applications [11], [176], [177].

Generally, the infrared image model can be formulated as follows [178]:

$$f_D(x,y) = f_T(x,y) + f_B(x,y) + f_N(x,y) \quad (2.25)$$

where f_D, f_T, f_B, f_N, and (x, y) are the original infrared image, the target image, the background image, the random noise image and the pixel location, respectively.

Small target shapes are assumed to be roughly circular without anisotropy and prevailing orientations [179]. Especially, a small target was modelled by a 2D Gaussian function in [180], [181]. That is,

$$s(x,y) = \gamma e^{-\frac{1}{2}\left((\frac{x}{\sigma_x})^2 + (\frac{y}{\sigma_y})^2\right)} \quad (2.26)$$

where γ, σ_x and σ_y separately determine the peak height, horizontal and vertical extent of the target. However, this assumption may not be valid in all cases. As shown in Fig.2.23(b) and (d), the small target may have a double-peak or "flat top" shapes

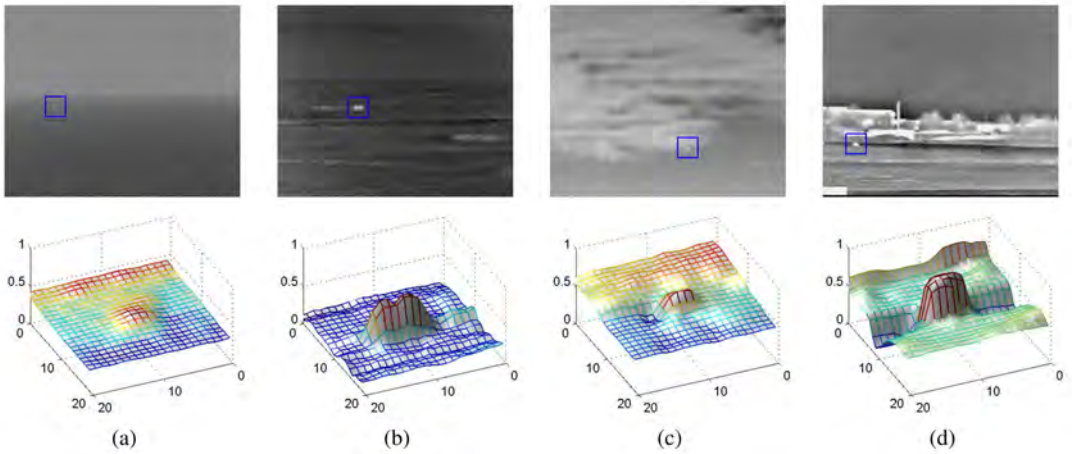


Figure 2.23: Representative targets (upper) and the corresponding 3-D surfaces (lower) in different backgrounds (normalized) [23] . (a) A dim small ship target in sea-sky background. (b) A bright ship target in sea-sky background. (c) A dim aeroplane target in sky cloud background. (d) A bright vehicle target in sky-ground background.

which can not be modelled as a single Gaussian function. The target size may vary from 2×2 to more than 10×10 pixels in infrared images according to its distance to the camera. (see Fig.2.23).

Most of infrared small target detection methods are always based on two assumptions: 1) backgrounds have the correlation in spatial domain and the stability in time domain, and occupy the low-frequency portion in the frequency domain; 2) small targets are unrelated to the background in spatial domain, and they dominate the high-frequency portion in frequency domain [182].

A number of single-frame and sequential detection methods have been proposed to detect small targets, where the former focus on estimating small targets in a single frame while the later make use of image sequences for target detection. These two type of detection methods are review as follows.

2.6.1 Single-frame Detection Methods

Single-frame detection methods can be roughly categorized into three groups depending on whether the focus is on the target image f_T , the background image f_B or both of them. Some detection methods focused on background image f_B and predicted the background image by various filters. Once the background image f_B is estimated, the target image f_T can be obtained by subtracting f_B from the original image f_D.

That is,

$$f_T = f_D - f_B. \tag{2.27}$$

In order to estimate background image f_B, Deshpande *et al.* made early attempt to propose max-mean and max-median filtering methods [183]; Gu *et al.* [178] developed a kernel-based nonparametric regression method; Bae [184] designed a spatial bilateral filter; Shang *et al.* [185] presented a line-based reconstruction method. Hadhoud *et al.* [186] proposed an adaptive filter, known as the two-dimensional least mean square filter (TDLMS), which produces a predicted background image by matching and tracking the desired image pixel after pixel. Based on the TDLMS, some new methods have been presented including two-dimensional block diagonal LMS adaptive filtering [187], improved 2D adaptive lattice algorithm [188], TDLMS filter based on neighborhood analysis [189]. Tom *et al.* [190], [191] developed a TopHat filter based on mathematical morphology where openings (erosion followed by dilation) and closings (dilation followed by erosion) operations were used to estimate background by removing target-like peaks. Along this line, a number of methods are derived to improve performance of the TopHat filter, including the toggle contrast operator [192], multiscale center-surround top-hat transform [193], hit-or-miss transform [194], adaptive morphological top-hat transform [195], and contour-structuring-element-based TopHat transformation [196]. The relationships of different modified top-hat transformations were analysed in [197].

A number of detection methods focus on the target image f_T and detect small targets by measuring the dissimilarity of the target region from its surrounding areas. In these methods, the luminance of the target region is always assumed to be higher than that of its surrounding areas. Based on this point, Wang et al. [198] formulated small target detection as a problem of finding extreme points in the infrared images and constructed the extremum filter by the cubic facet model. Similarly, an extremum filter generated using least squares support vector machine (LS-SVM) was developed [199]. Chen *et al.* [200] present a local contrast measure to calculate the gray difference between the target region and its neighborhoods. Han *et al.* [201] subsequently developed a refinement of the local contrast measure to improve dark target

detection performance. Deng *et al.* [202] designed a weighted local contrast measure to suppress cloudy-sky backgrounds. Wei *et al.* [203] proposed a multiscale patch-based contrast measure to improve detection performances. Kim [204] developed a spatial filter called the min-local-LoG filter to detect blob structures by decomposing the 2D Laplacian of Gaussian (LOG) filter into several 1D filters.

Some detection methods jointly consider all three components, namely f_B, f_T and f_N, in the low-rank framework [23], [205], [206]. Using local patch construction, f_B, f_T and f_N are transformed into corresponding patch-images where the target patch-image and background patch-image are assumed to be a sparse matrix and a low-rank matrix, respectively. Then the small target detection task is transformed into an optimization problem recovering the low-rank and sparse matrices.

2.6.2 Sequential Detection Methods

Sequential detection methods can be roughly categorized into two groups including detection before track (DBT) and track before detection (TBD). The DBT methods first detect potential targets in single frame, then exploit the continuity of targets trajectories to reject false positives. Thus the performance of the DBT methods greatly depends on detection results from the single frame. In contrast, the TBD methods always seek potential target trajectories before detection, then accumulate the signal energy of the target along the trajectory to enhance target signal energy and eliminate false positives simultaneously.

One of the early classical TBD methods is 3D directional filtering (3DDF) which is also called matched filtering [207]. The 3DDF first transforms entire image sequence into frequency domain via three-dimensional Fourier transformation. Then a matched filter which is designed using prior information about small targets velocity, is applied to the transformed image for detecting targets with the same velocity. That is, the 3DDF can only detect the known constant-velocity moving targets. To overcome this issue, Porat and Friedlander [208] used a bank of the 3DDFs for all possible target motion directions. They treated all possible target directions separately and derived a decision rule about the existence of targets in each candidate direction. To further

improve the ability of detecting weaker targets, a 3D double directional filter (3DDDF) was presented [209] by increasing the target energy accumulation ability. To increase the speed of filtering, Liu and Zuo [210] developed an improved 3D directional filtering method which employs a dual diffusion partial differential equation (DFPDE) to pre-whiten each image and then uses a wide-to-exact search technique to quickly determine small target region.

Different from the above methods considering the small target detection problem in a 3D spatial-temporal space, some TBD methods formulate this problem in the context of a hypothesis testing procedure on individual pixel temporal profiles. Tzannes *et al.* [211] assumed that the temporal behavior of the background noise, target, and clutter all have reasonable deterministic and statistical models on a single pixel basis. Then small targets are detected by comparing the temporal profile of each pixel with the derived model using statistic hypothesis method. Similarly, Lim *et al.* [212] used the variance of the temporal profile as a criterion to discriminate small targets from background. Liu *et al.* [213] developed an algorithm employing the connecting line of the stagnation points (CLSP) of the temporal profile as the baseline. The deviation of the temporal profile and its CLSP is calculated and analysed for small target detection. Bae [184] combined the temporal profile with spatial bilateral filter (BF) to detect small targets where the BF is operated as the background predictor by blurring the target region through the adjustment of standard deviation.

The DBT methods first detect small targets and suppress background clutter in a single frame, then use multiframe accumulation, autocorrelation or other techniques to filter out random noise or false alarms. The performance of the DBT methods largely depend on the detection result of each frame. Wang *et al.* [214] presented an infrared small target detection method based on support vector machines (SVM) in the wavelet domain. The feature image is extracted by the SVM in wavelet domain for each frame, then fused in consecutive frames to enhance small target signals. Zhang *et al.* [215] developed an adaptive anisotropic filter based on a modified partial differential equation (AFMPDE). In each frame, the structured background is adaptively eliminated and the target signal is enhanced by a regularizing operator.

2.7 Chapter Summary

We have reviewed three motion-sensitive neural models, traditional motion detection and infrared small target detection approaches in the previous sections. Small target motion detection aims to detect objects of interest that move against cluttered natural environments and appear as small dim speckles in images. It means not only locating small moving targets, but also distinguishing them from large moving objects.

Detecting moving objects against natural backgrounds is relatively simple, but discriminating them with regard to their sizes is much more challenging. The traditional motion detection methods, the LGMD-based and the LPTC-based models all perform well in motion detection, nevertheless they do not show size selectivity and are powerless for discriminating small moving targets. A number of infrared-based methods have been proposed to detect small target motion in infrared images. However, their detection performances strongly rely on significant temperature differences between the background and objects of interest which are rare in the real world.

The STMD-based and the FD-based models are developed for small target motion detection in natural images. They demonstrate strong size selectivity and can effectively inhibit large moving objects. However, their methods to produce size selectivity are different. Specifically, the size selectivity of the STMD-based models is shaped by a lateral inhibition mechanism, whereas that of the FD-based models comes from the inhibition by the background motion which is strongly dependent on relative motion between moving objects and the background. That is, FD-based models would be powerless in the case the moving object and the background are relatively static.

The following chapters focus on building quantitative STMD-based models to provide robust and economic solutions of small target detection for an artificial vision system. The proposed models satisfying the basic properties of the STMD neurons are able to detect small moving targets against cluttered backgrounds, but also discriminate them from large moving objects and small-target-like background features.

Chapter 3

DSTMD – a STMD with Directional Selectivity

Intelligent robots have shown great potential in reshaping human life in the future. However, artificial visual systems so far are still struggling to provide a robot with the required capacity to respond to the dynamic visual world in real time, like many animal species do. Among many visual functionalities, detecting small moving targets is one of the most important abilities for many animal species, e.g., finding mates in the distance, and it is also critical for a robot to track small targets in a cluttered background.

Small target motion detection against cluttered moving backgrounds is always a challenge for artificial visual systems due to limited physical cues of small targets, free motion of camera, and extremely cluttered backgrounds. In the natural world, this challenge has been elegantly solved by many insect species, as evidenced by their high successful rate for pursuing mates or prey. The exquisite selectivity of insects for small targets comes from a class of specific neurons called small target motion detectors (STMDs) [17], [28], [29]. In order to provide solutions of small target motion detection for artificial visual systems, plenty of efforts have been made to develop STMD-based visual neural model, including ESTMD [34] and the hybrid models [35]. Although these existing models are sensitive to small target motion, they did not systematically study direction selectivity of the STMD neurons which contributes to motion direction estimation. Computationally modelling these directionally selective STMD neurons

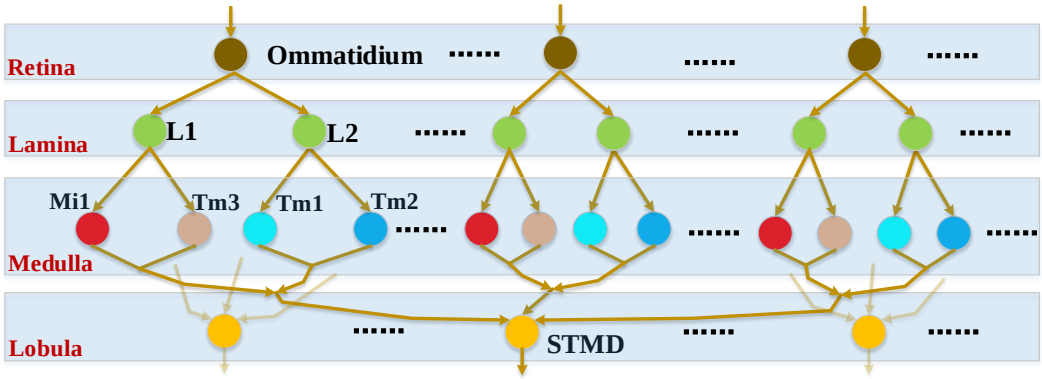


Figure 3.1: Wiring sketch of the insect's visual system. The insect's visual system consists of four neural layers, including retina, lamina, medulla and lobula (from top to bottom). Each neural layer contains numerous specialized neurons illustrated by coloured circular nodes. Luminance signals are firstly perceived by ommatidia, further processed by large monopolar cells (LMC, i.e., L1 and L2) and medulla neurons (Mi1, Tm1, Tm2, Tm3), finally integrated in STMD neurons. Note that the connection between the four medulla neurons and the STMD neuron is speculative.

would not only deepen our understanding of the biological visual processing, but also endow artificial visual systems with capacity to track and intercept small targets.

This chapter develops a neural model to simulate the specific STMD neurons with direction selectivity named as DSTMD. It can detect small moving targets against cluttered background, but also estimate their motion directions. To introduce direction selectivity, a new correlation mechanism is proposed via correlating luminance signals from two different pixels. Then, a lateral inhibition mechanism acting on correlation outputs is used for size selectivity. Finally, a population vector algorithm is devised to encode motion directions of small targets. Systematic experiments are carried out to validate the developed neural model in complex environments.

The remainder of this chapter is organized as follows. Section 3.1 describes the proposed neural model in details. In Section 3.2, experiments are carried out to test the performances of the proposed neural model. We give further discussions in Section 3.3 and finally in Section 3.4, we conclude this chapter.

3.1 Formulation of the Model

Following the typical multi-stage view of motion detection in the insect's visual system (schematically illustrated in Fig. 3.1), we devised a directionally selective

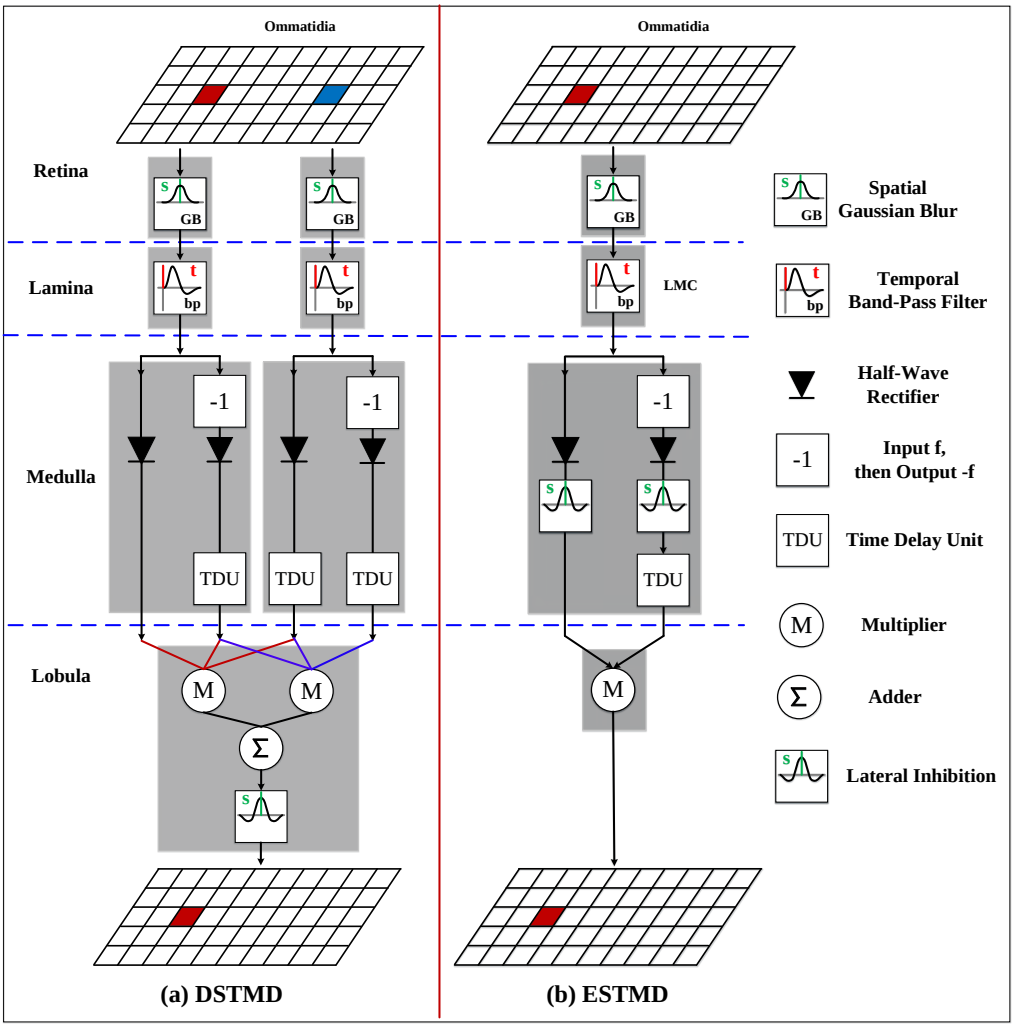


Figure 3.2: Schematic illustration of the proposed DSTMD and the existing ESTMD models, both of which exhibit selectivity for dark small targets. The DSTMD integrates signals from two different positions (the red and blue pixels in the retina), whereas the ESTMD utilizes signals from a single position. Since one of the two positions (the blue pixel) has multi alternatives in the image plane corresponding to different preferred directions, the DSTMD can produce multi directionally selective outputs at each position of the lobula. However, the ESTMD just has a single output without direction selectivity.

STMD-based neural model **(DSTMD)** in this chapter. Fig. 3.2(a) shows the schematic diagram of one DSTMD cell and its presynaptic neural network, whereas Fig. 3.2(b) presents the existing ESTMD model for comparison. As can be seen, the proposed neural model is composed of four neural layers including the retina, lamina, medulla and lobula. These four sequentially arranged neural layers have specific functions and cooperate together for small target motion detection. Fig. 3.3 shows the outputs of the four neural layers in the DSTMD. As can be seen, luminance signals from pixel A and B are first received and smoothed by the retina layer, then propagated to the lamina

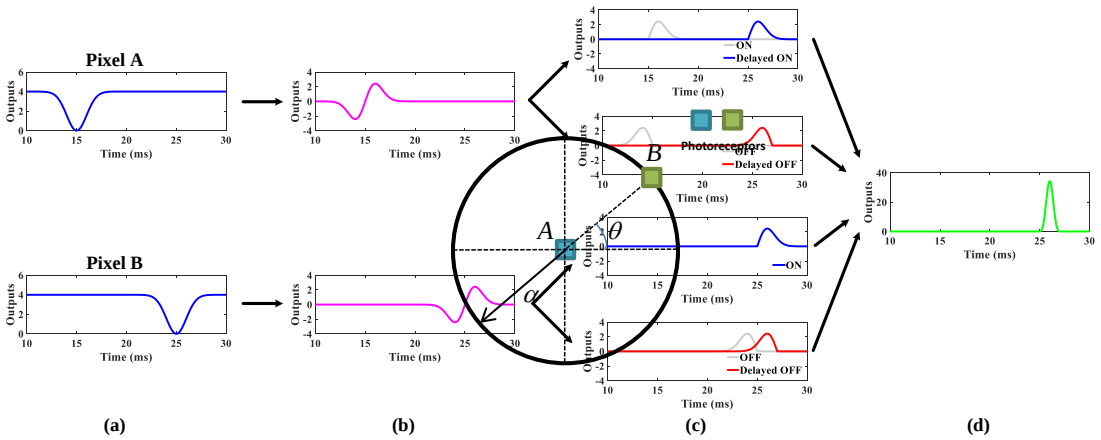


Figure 3.3: The outputs of (a) the retina layer, (b) the lamina layer, (c) the medulla layer, and (d) the lobula layer.

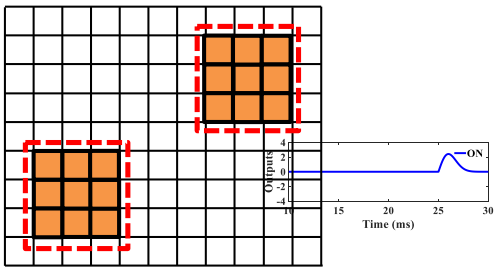


Figure 3.4: Schematic illustration of the mapping from pixels to photoreceptors. Each small square denotes a pixel, corresponding to a photoreceptor. Each red dotted rectangle which contains multiple pixels (photoreceptors), represents the visual region of an ommatidium.

layer to extract luminance changes with respect to time. The medulla layer separates the extracted luminance change signals into separated into luminance increase (ON) and decrease components (OFF), then further aligns them in the time domain by appropriate time delays. To produce a large response for the small moving target, these temporally aligned ON and OFF components are recombined in the lobula layer by a multiplication and lateral inhibition. In the following subsections, we will elaborate on the function of each neural layer and provide mathematical descriptions.

3.1.1 Retina Layer

In the insect's visual system, the retina layer contains a great number of ommatidia (see Fig. 3.1). Each ommatidium is composed of eight photoreceptors. Each photoreceptor views a small region of the whole visual filed and supplies a 'pixel' of luminance information to ommatidia [216].

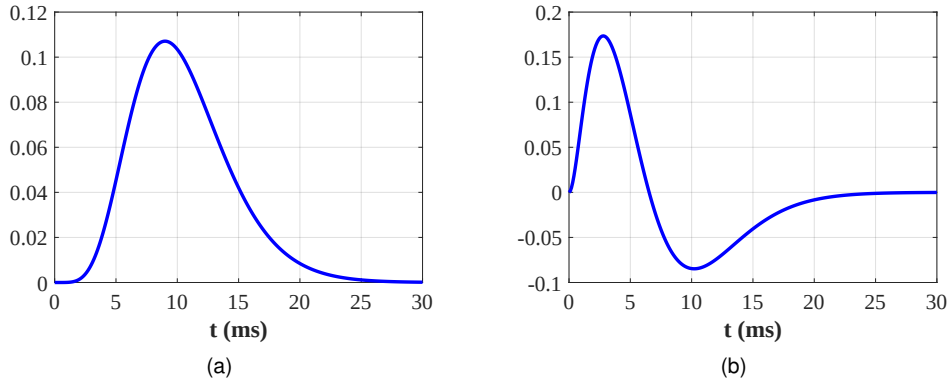


Figure 3.5: (a) Gamma kernel $\Gamma_{n,\tau}(t)$ where $n = 6$, $\tau = 9$. (b) Impulse response of temporal band-pass filter $H(t)$ where $n_1 = 2$, $\tau_1 = 3$, $n_2 = 6$, $\tau_2 = 9$.

Image sequences are inputs to the developed neural model, so a mapping from pixels to photoreceptors needs to be constructed first. As depicted in Fig. 3.4, each small square denotes a pixel, corresponding to a photoreceptor. The red dotted rectangle which contains multiple pixels (photoreceptors), represents the visual region of an ommatidium. Specifically, let $I(x, y, t) \in \mathbb{R}$ denote varying luminance values captured by photoreceptors where x, y and t are spatial and temporal field positions. Each ommatidium is modelled by a Gaussian filter to smooth the input luminance signals (see Fig. 3.2). That is, the output of an ommatidium $P(x, y, t)$ with viusal region centered at (x, y) is given by,

$$P(x, y, t) = \iint I(u, v, t) G_{\sigma_1}(x - u, y - v) du dv \tag{3.1}$$

where $G_{\sigma_1}(x, y)$ is a Gaussian function, defined as

$$G_{\sigma_1}(x, y) = \frac{1}{2\pi\sigma_1^2} \exp(-\frac{x^2 + y^2}{2\sigma_1^2}). \tag{3.2}$$

3.1.2 Lamina Layer

In the insect's visual system, large monopolar cells (LMCs), such as L1 and L2, are located in the lamina layer and receive the smoothed luminance signals from the ommatidia [7], as shown in Fig. 3.1. The LMCs have been shown to be sensitive to luminance changes with respect to time which are induced by object motion. More precisely, they exhibit positive responses to luminance increments while negative re-

sponses to luminance decrements [7], [217].

To extract luminance changes induced by object motion, each LMC is modelled as a temporal band-pass filter in the developed neural model. The impulse response of the band-pass filter denoted by $H(t)$, is given by the difference of two Gamma kernels $\Gamma_{n,\tau}(t)$ which have been widely used in studying neural processing of time varying patterns [218]–[220]. That is,

$$H(t) = \Gamma_{n_1,\tau_1}(t) - \Gamma_{n_2,\tau_2}(t) \tag{3.3}$$

$$\Gamma_{n,\tau}(t) = (nt)^n \frac{\exp(-nt/\tau)}{(n-1)!\tau^{n+1}} \tag{3.4}$$

where n and τ refers to the order and time constant of the Gamma kernel $\Gamma_{n,\tau}(t)$, respectively. The illustration of $\Gamma_{n,\tau}(t)$ and $H(t)$ is presented in Fig. 3.5. Let $L(x, y, t)$ denote the output of a LMC located at (x, y). Then $L(x, y, t)$ can be calculated by convolving the output of ommatidia $P(x, y, t)$ with $H(t)$,

$$L(x, y, t) = \int P(x, y, s)H(t - s)ds. \tag{3.5}$$

The output of LMCs $L(x, y, t)$ is able to reflect luminance changes of pixel (x, y) over time. Specifically, a positive $L(x, y, t)$ means luminance increase while a negative $L(x, y, t)$ suggests luminance decrease.

3.1.3 Medulla Layer

In the insect's visual system, medulla neurons including Tm1, Tm2, Tm3 and Mi1, are downstream neurons of the LMCs (see Fig. 3.1). These four neurons constitute four parallel channels to process the output of the LMCs. Specifically, the Mi1 and Tm3 respond selectively to luminance increments, with the response of the Mi1 delayed relative to the Tm3 [7]. Conversely, the Tm1 and Tm2 respond selectively to luminance decrements, with the response of the Tm1 delayed relative to the Tm2 [63].

The DSTMD and the existing ESTMD [34] adopt two different methods to model these four medulla neurons. These two modelling methods are described as follows, respectively.

1) Medulla Neuron Modeling of DSTMD: Since the Tm3 and Tm2 neurons respond strongly to luminance increments and decrements, the DSTMD models them as half-wave rectifiers to split the output of LMCs $L(x, y, t)$ into positive and negative parts. That is,

$$S^{\text{Tm3}}(x, y, t) = [L(x, y, t)]^{+} \tag{3.6}$$

$$S^{\text{Tm2}}(x, y, t) = [-L(x, y, t)]^{+} \tag{3.7}$$

where $[x]^{+}$ represents $\max(x, 0)$. The Mi1 and Tm1 are temporally delayed versions of the Tm3 and Tm2, respectively, so the output of the Tm3 and Tm2 are convolved with a Gamma kernel to implement time delay. Then the outputs of the Mi1 and Tm1 $S^{\text{Mi1}}_{(n,\tau)}(x, y, t)$, $S^{\text{Tm1}}_{(n,\tau)}(x, y, t)$ is given by

$$S^{\text{Mi1}}_{(n,\tau)}(x, y, t) = \int S^{\text{Tm3}}(x, y, s)\Gamma_{n,\tau}(t - s)ds \tag{3.8}$$

$$S^{\text{Tm1}}_{(n,\tau)}(x, y, t) = \int S^{\text{Tm2}}(x, y, s)\Gamma_{n,\tau}(t - s)ds \tag{3.9}$$

where n, τ stand for the order and time constant of the Gamma kernel, determining the order and time-delay length of the time delay unit (TDU) (see Fig. 3.2).

2) Medulla Neuron Modeling of ESTMD: Comparing with the DSTMD, the ESTMD implements a lateral inhibition mechanism on the outputs of medulla neurons to suppress luminance changes caused by large moving objects. As shown in Fig. 3.2(b), the lateral inhibition mechanism is applied after the half-wave rectifiers. Thus, the outputs of the Tm3 and Tm2 denoted by $\tilde{S}^{Tm3}$ and $\tilde{S}^{Tm2}$, are give by,

$$\tilde{S}^{\text{Tm3}}(x, y, t) = \iint [L(u, v, t)]^{+} \cdot W_s(x - u, y - v)dudv \tag{3.10}$$

$$\tilde{S}^{\text{Tm2}}(x, y, t) = \iint [-L(u, v, t)]^{+} \cdot W_s(x - u, y - v)dudv \tag{3.11}$$

where $W_s(x, y)$ is the lateral inhibition kernel, defined as

$$W_s(x, y) = A[g(x, y)]^{+} + B[g(x, y)]^{-} \tag{3.12}$$

$$g(x, y) = G_{\sigma_2}(x, y) - e \cdot G_{\sigma_3}(x, y) - \rho \tag{3.13}$$

where A, B, e, ρ are constant and $G_\sigma(x, y)$ is a Gaussian function. The outputs of the Tm3 and Tm2 are further temporally delayed to obtain the outputs of the Mi1 and Tm1. That is,

$$\tilde{S}^{\text{Mi1}}_{(n,\tau)}(x, y, t) = \int \tilde{S}^{\text{Tm3}}(x, y, s) \cdot \Gamma_{n,\tau}(t - s)ds \tag{3.14}$$

$$\tilde{S}^{\text{Tm1}}_{(n,\tau)}(x, y, t) = \int \tilde{S}^{\text{Tm2}}(x, y, s) \cdot \Gamma_{n,\tau}(t - s)ds \tag{3.15}$$

where $\tilde{S}^{\text{Mi1}}_{(n,\tau)}$ and $\tilde{S}^{\text{Tm1}}_{(n,\tau)}$ denote the outputs of the Mi1 and Tm1, respectively.

Here, we discuss the implementation of the lateral inhibition mechanism. Existing biological research [14] asserts that the size selectivity of the STMD neurons is shaped by a lateral inhibition mechanism. However, where this lateral inhibition mechanism occurs remains elusive. Although the ESTMD implements this lateral inhibition mechanism on medulla neurons, it is just speculative and there is no neuroanatomical evidence for it. On the other hand, we notice that the lobula plate tangential cells (LPTCs) also receive signals from medulla neurons [7], [221]. If the medulla neurons which provide signals to the LPTC neurons, are laterally inhibited, the LPTC neurons would show strong size selectivity (will be demonstrated in the experiment Section 3.2.2). This may conflict with the finding that the LPTCs do not exhibit size selectivity [222]–[224]. To satisfy both size selectivity of the STMDs and size insensitivity of the LPTCs, we infer that this lateral inhibition mechanism may happen on the STMD pathway rather than medulla neurons in the implementation of our proposed neural model.

3.1.4 Lobula Layer

In the insect's visual system, the STMD neurons integrate signals from the medulla layer. They respond strongly to small target motion, but show weak or no response to wide-field motion (size selectivity) [17], [28], [29]. Besides, some STMDs exhibit strong responses to small target motion oriented along a preferred direction, but show weak or no response to opposite-direction motion (direction selectivity) [13], [30].

In the proposed DSTMD, a new correlation mechanism and a lateral inhibition

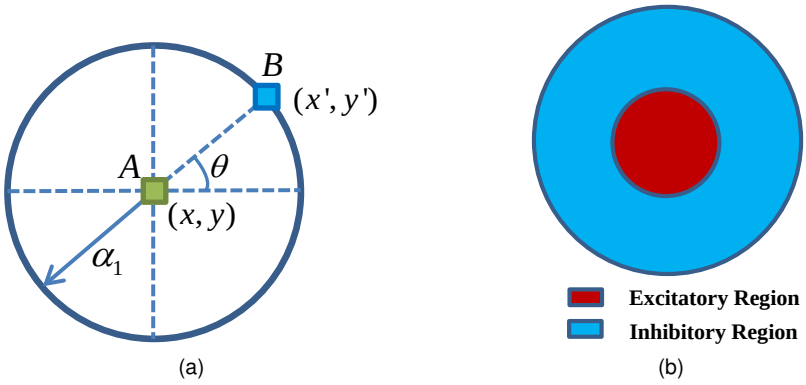


Figure 3.6: (a) Schematic illustration of relative position between A (x, y) and B (x', y'). α_1 is the distance between A and B while θ is the angle between line segment AB and the horizontal line. (b) Schematic illustration of excitatory and inhibitory regions of the lateral inhibition mechanism.

mechanism are introduced for direction and size selectivities, respectively. For comparison with the DSTMD, the existing non-directionally selective ESTMD [34] is also presented as follow.

1) ESTMD: To produce larger responses to small moving targets, the ESTMD multiplies the luminance increase signal $\tilde{S}^{\text{Tm3}}(x, y, t)$ with temporally delayed luminance decrease signal $\tilde{S}^{\text{Tm1}}_{(n_3,\tau_3)}(x, y, t)$ at each pixel (x, y). The output of a STMD neuron located at (x, y), denoted by $\tilde{D}(x, y, t)$, is given by,

$$\tilde{D}(x, y, t) = \tilde{S}^{\text{Tm3}}(x, y, t) \times \tilde{S}^{\text{Tm1}}_{(n_3,\tau_3)}(x, y, t). \tag{3.16}$$

As can be seen from (3.16), the output of a STMD neuron located at (x, y) is determined by medulla signals at the same pixel. However, medulla signals from at least two different pixels are needed for estimating motion direction [131], so the ESTMD is able to detect the presence of target motion, but not motion directions.

2) DSTMD: To introduce direction selectivity and produce large responses to moving object, the DSTMD multiplies the outputs of medulla neurons located at two different pixels together, i.e., (x, y) and $(x'(\theta), y'(\theta))$ which are defined as

$$\begin{aligned} x'(\theta) &= x + \alpha \cos\theta \\ y'(\theta) &= y + \alpha \sin\theta \end{aligned} \tag{3.17}$$

where α is a constant and θ denotes the preferred direction of the STMD neuron. The output of a STMD neuron located at (x, y) with a preferred direction θ, denoted by $D(x, y, t, \theta)$, is defined as,

$$\begin{aligned} D(x,y,t,\theta) = S^{\text{Tm3}}(x,y,t) \times \Big\{ & S^{\text{Tm1}}_{(n_5,\tau_5)}(x,y,t) \\ & + S^{\text{Mi1}}_{(n_4,\tau_4)}(x'(\theta),y'(\theta),t) \Big\} \times S^{\text{Tm1}}_{(n_6,\tau_6)}(x'(\theta),y'(\theta),t). \end{aligned} \tag{3.18}$$

As shown in (3.18), four medulla neural signals from two different positions, i.e., (x, y) and (x', y'), are used to define the output of a STMD neuron located at (x, y) (see Fig. 3.2, two multipliers and one adder). The schematic illustration of relative position between (x, y) and (x', y') is presented in Fig. 3.6(a). For a given position (x, y), we can choose a series of (x', y') corresponding to different preferred directions θ. Utilizing (3.18), a series of correlation outputs $D(x, y, t, \theta)$ with different preferred directions θ can be defined. For a given direction θ, $D(x, y, t, \theta)$ exhibits the strongest output to small target motion oriented along direction θ, with weak or no outputs to motion oriented along other directions. That is, the output of STMD neurons $D(x, y, t, \theta)$ shows direction selectivity.

After the medulla signal multiplication, the obtained $D(x, y, t, \theta)$ responds strongly to both large and small moving objects. To suppress responses to large moving objects, the $D(x, y, t, \theta)$ is laterally inhibited by convolving with an inhibition kernel $W_s(x, y)$,

$$D_I(x,y,t;\theta) = \iint D(u,v,t,\theta) W_s(x-u, y-v) du dv \tag{3.19}$$

where $D_I(x, y, t; \theta)$ denotes the signal after lateral inhibition and the inhibition kernel $W_s(x, y)$ is defined in (3.12). Fig. 3.6(b) displays the schematic illustration of the inhibition kernel $W_s(x, y)$. As can be seen, the $W_s(x, y)$ contains two components, i.e., excitatory and inhibitory regions. Generally, the surround inhibition of kernel $W_s(x, y)$ is set as three times as large as the center excitation. Due to this mechanism, a target will receive strong inhibition if its size exceeds the excitatory region; In contrast, when the target is smaller than the excitatory region, the amount of excitation will increase as the rise of target size. That is, the DSTMD prefers the target whose size is equal to

the excitatory region and exhibits size selectivity.

Following the lateral inhibition mechanism implemented on spatial field for size selectivity, the DSTMD further inhibits model output $D_I(x, y, t, \theta)$ at directions more than $45°$ apart by convolving $D_I(x, y, t, \theta)$ with an directional inhibition kernel $W_d(\theta)$,

$$E(x, y, t, \theta) = \int D_I(x, y, t, \varphi) W_d(\theta - \varphi) d\varphi \tag{3.20}$$

$$W_d(\theta) = G_{\sigma_4}(\theta) - G_{\sigma_5}(\theta) \tag{3.21}$$

where $G_\sigma(x, y)$ is a Gaussian function and $E(x, y, t, \theta)$ is regarded as the final output of the STMD neurons. Through this directional lateral inhibition, the developed DSTMD is able to avoid producing strong responses along two completely opposite directions.

3.1.5 Motion Direction Estimation

In the insect's visual system, the STMD neurons are believed to be upstream of target selective descending neurons (TSDNs) [28], [30], [31]. Further biological research [31] found that eight pairs of the TSDNs are able to encode motion direction of targets by a population vector algorithm.

The proposed neural model estimates motion direction of a small target by populating the model output $E(x, y, t, \theta)$ along different directions θ,

$$MD(t) = \sum_{(x,y)\in\Omega} \sum_{\theta} (E(x, y, t, \theta) \cos\theta, E(x, y, t, \theta) \sin\theta) \tag{3.22}$$

where $MD(t)$ denotes the motion direction of the small target at time t; $(x, y) \in \Omega$ represents the position of the STMD neurons which respond to the small moving target; $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}\}$.

3.1.6 Parameter Setting

Parameters of the developed neural model (DSTMD) and ESTMD are given in Table 3.1. These parameters are tuned manually and will not be changed in the following experiments unless stated.

Table 3.1: Parameters of the DSTMD and ESTMD

Equation	Parameters
(3.1)	$\sigma_1 = 1$
(3.3)	$n_1 = 2, \tau_1 = 3, n_2 = 6, \tau_2 = 9$
(3.12)	$A = 1, B = 3$
(3.13)	$\sigma_2 = 1.5, \sigma_3 = 3, e = 1, \rho = 0$
(3.16)	$n_3 = 5, \tau_3 = 25$
(3.17)	$\alpha = 3$
(3.18)	$n_4 = 3, \tau_4 = 15, n_5 = 5, \tau_5 = 25, n_6 = 8, \tau_6 = 40$
(3.21)	$\sigma_4 = 1.5, \sigma_5 = 3$

The developed neural model is written in Matlab (The MathWorks, Inc., Natick, MA). The computer used in the experiments is a PC with one 2.50 Ghz CPU (Core i7 4710MQ) and windows 7 operating system. The source code can be found at https://github.com/wanghongxin/DSTMD.

3.2 Results and Discussions

The proposed neural model is tested on image sequences produced by Vision Egg [225], which is available at http://visionegg.org. The Vision Egg is an open-source programming library that allows scientists to produce arbitrary visual stimuli. Such stimuli involve traditional stimuli such as sinusoidal gratings, or may be more complex, 3D, and naturalistic scenes. The image sequences used in the experiment can be divided into two categories depending on background types. The first category contains image sequences showing small target motion against white backgrounds. This category is used to test basic properties of the proposed neural model, such as tuning properties (see subsection B and C), direction selectivity (see subsection D). The other category contains image sequences showing small target motion against naturally cluttered backgrounds. This category is utilized to evaluate the detection performance of the proposed neural model in complex backgrounds (see subsection A and E). All image sequences can be reproduced by the Vision Egg with the same parameters given before each experiment. The video images are 500 (in horizontal) by 250 (in vertical) pixels and temporal sampling frequency is set as 1000 Hz.

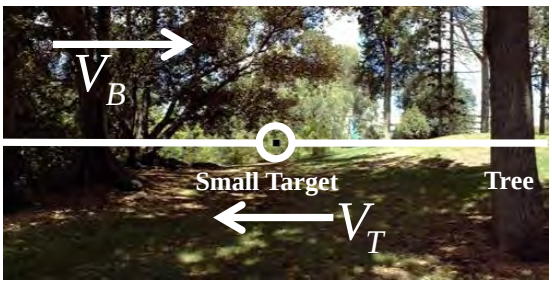


Figure 3.7: Representative frame at time $t_0 = 1000$ ms whose resolution is 500 pixels (in horizontal) by 250 pixels (in vertical). The small target (the black block) whose size and luminance are set as 5×5 pixels and 0, is moving against the cluttered background. The velocities of the small target and background are all equal to 250 pixel/s, and arrow V_T and V_B denote their motion directions, respectively. The tree which is regarded as a large object, is also moving due to the background motion.

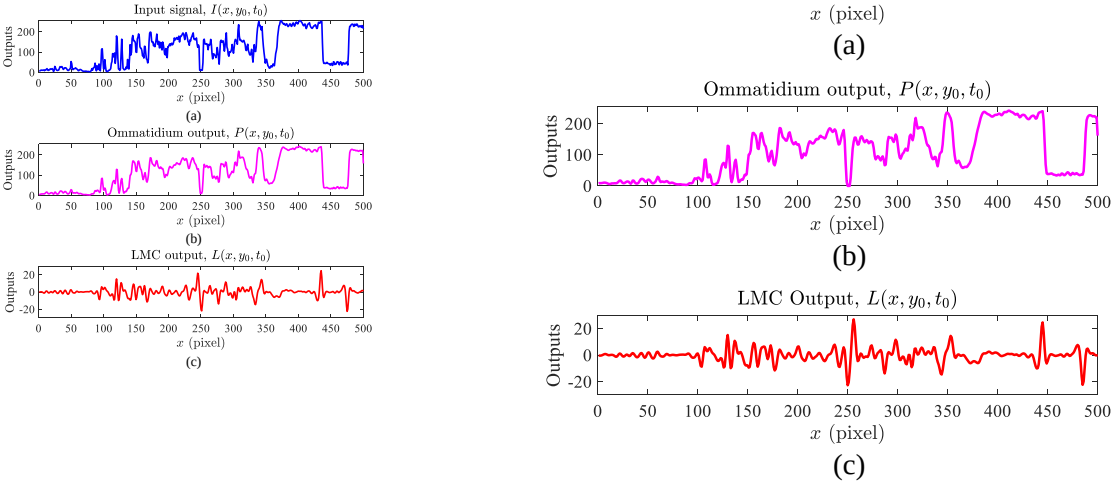


Figure 3.8: In each subplot, the horizontal axis denotes x coordinate while the vertical axis represents neural outputs. (a) Input luminance signal $I(x, y_0, t_0)$. (b) Ommatidium output $P(x, y_0, t_0)$. (c) LMC output $L(x, y_0, t_0)$.

3.2.1 Contribution of Various Neurons

To evaluate characteristics of various neurons in the developed neural model, we observe and analyze their outputs at each layer. For an input image sequence $I(x, y, t)$, where $x \in [0, 500]$ pixel, $y \in [0, 250]$ pixel, $t \in [0, 1000]$ ms (see Fig. 3.7), we separately fix y and t as $y_0 = 125$ pixel and $t_0 = 1000$ ms, then illustrate $I(x, y_0, t_0)$ with regard to x in Fig. 3.8(a). Similarly, the outputs of other neurons with respect to x are presented in the subplots below.

Fig. 3.8(a)-(c) shows the input luminance signal $I(x, y_0, t_0)$, ommatidium output $P(x, y_0, t_0)$ and LMC output $L(x, y_0, t_0)$, respectively. Compared to the input signal,

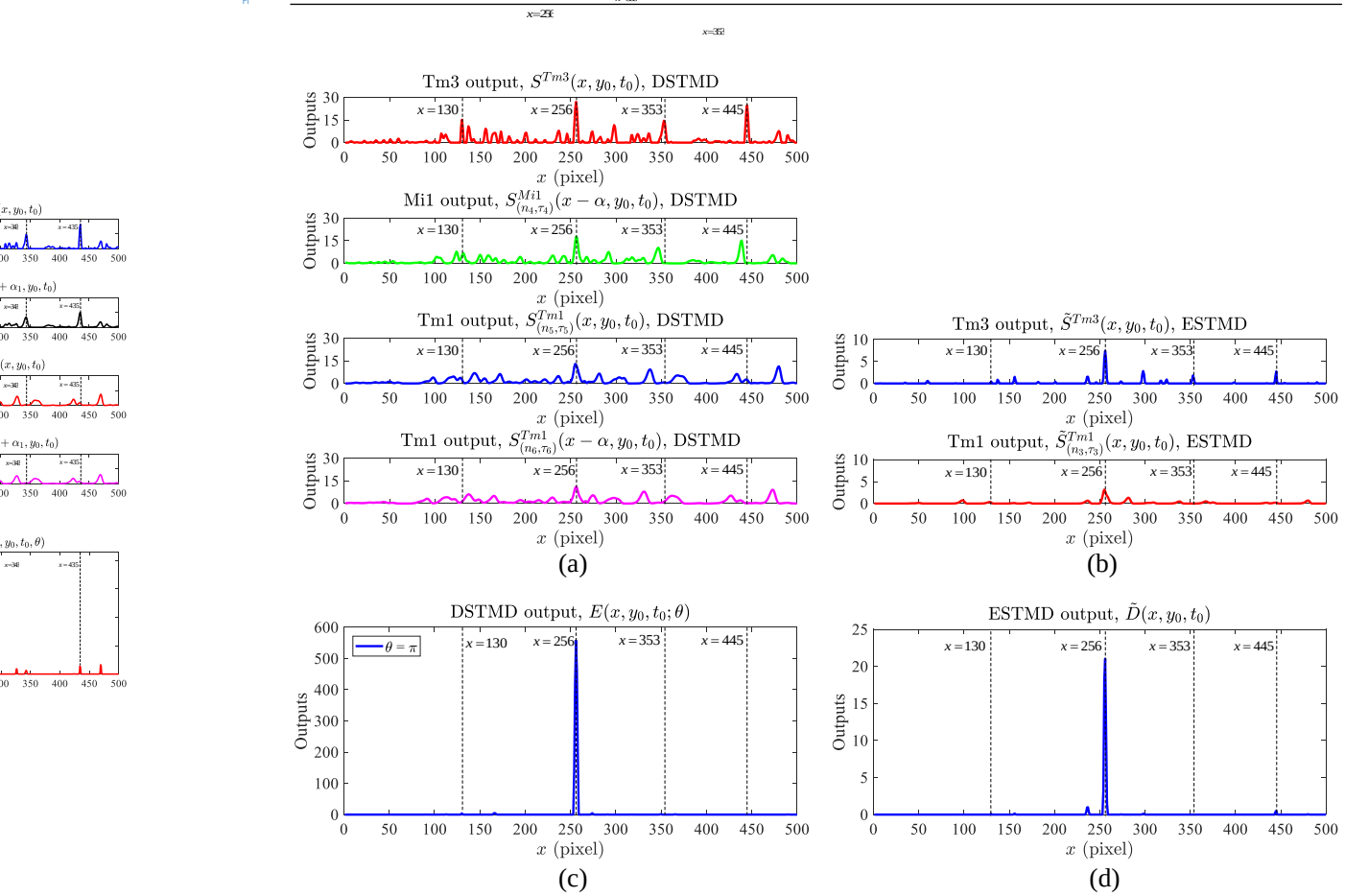


Figure 3.9: In each subplot, the horizontal axis denotes x coordinate while the vertical axis represents neural outputs. (a) Four inputs of the DSTMD when the preferred direction θ is set as π, i.e. $S^{\text{Tm3}}(x, y_0, t_0)$, $S^{\text{Mi1}}_{(n_4,\tau_4)}(x-\alpha, y_0, t_0)$, $S^{\text{Tm1}}_{(n_5,\tau_5)}(x, y_0, t_0)$ and $S^{\text{Tm1}}_{(n_6,\tau_6)}(x-\alpha, y_0, t_0)$. (b) Two inputs of the ESTMD, i.e. $\tilde{S}^{\text{Tm3}}(x, y_0, t_0)$ and $\tilde{S}^{\text{Tm1}}_{(n_3,\tau_3)}(x, y_0, t_0)$. (c) DSTMD output $E(x, y_0, t_0, \theta)$ when the preferred direction θ is equal to 0. (d) ESTMD output $\tilde{D}(x, y_0, t_0)$.

the ommatidium output demonstrates little difference and is just slightly smoothed. This is because the ommatidium is modelled as a spatial Gaussian filter to smooth the input luminance signals. The LMC output displays significant difference from the ommatidium output. More precisely, the LMC output becomes the band-pass-filtered version of the ommatidium output. From the other perspective, the LMC output reveals the luminance changes of pixels, where the positive values correspond to luminance increase while the negative values suggest luminance decrease.

Fig. 3.9(a) demonstrates the four medulla signals used for the multiplication in the DSTMD when its preferred direction θ is set as π. In addition to the luminance increase signal $S^{\text{Tm3}}(x, y_0, t_0)$ and delayed luminance decrease signal $S^{\text{Tm1}}(x, y_0, t_0)$ at pixel (x, y_0), the DSTMD uses the delayed luminance increase signal $S^{\text{Mi1}}_{(n_4,\tau_4)}(x, y_0, t_0)$ and delayed luminance decrease signal $S^{\text{Tm1}}(x, y_0, t_0)$ at pixel $(x-\alpha, y_0)$ to estimate object

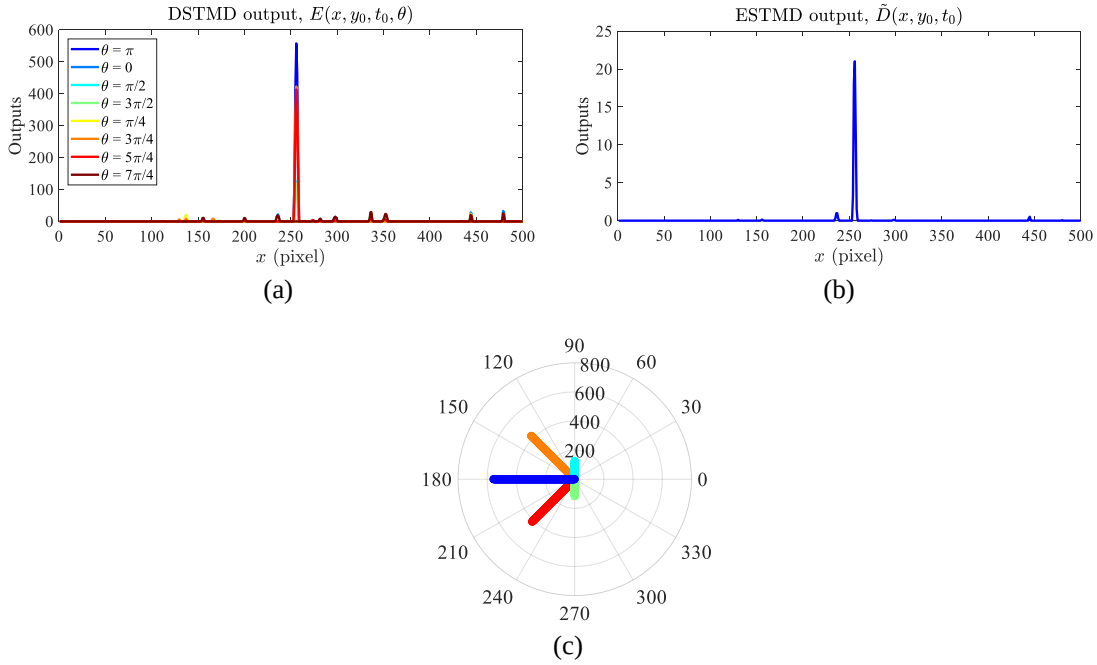


Figure 3.10: (a) DSTMD outputs $E(x, y_0, t_0, \theta)$ along eight preferred directions $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}\}$. (b) ESTMD output $\tilde{D}(x, y_0, t_0)$ without direction selectivity. (c) DSTMD outputs at position $x = 256$ along eight preferred directions. In the polar coordinate system, the angular coordinate represents the preferred direction θ while the radial coordinate denotes the STMD output.

motion along direction π. However, the ESTMD only utilizes the luminance increase signal and delayed luminance decrease signal at pixel (x, y_0), i.e., $\tilde{S}^{\text{Tm3}}(x, y_0, t_0)$ and $\tilde{S}^{\text{Tm1}}_{(n_3,\tau_3)}(x, y_0, t_0)$, which are unable to introduce direction information (see Fig. 3.9(b)).

Fig. 3.9(c) displays the output of the DSTMD when the preferred direction θ is equal to π, while Fig. 3.9(d) presents the output of the ESTMD. As it is shown, both the DSTMD and ESTMD exhibit high responses at position of the small target $x = 256$ where the luminance increase and decrease signals induced by the small target are aligned together via time delay (see Fig. 3.9(a) and (b)). At the other positions such as the position of the tree $x = 445$, the luminance increase and decrease signals show low aligning probability, so model outputs are much weak or even close to 0.

Note that Fig. 3.9(c) only illustrates the DSTMD output at direction π. To estimate motion direction, the outputs of the DSTMD along eight preferred directions are calculated and presented in Fig. 3.10(a). In contrast, the ESTMD only has one output $\tilde{D}(x, y_0, t_0)$ lacking of direction information, as shown in Fig. 3.10(b). To clearly demonstrate direction selectivity, the DSTMD outputs to the small target at pixel $x = 256$ are presented in polar coordinate (see Fig. 3.10(c)). As can be seen, the

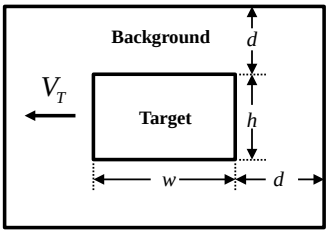


Figure 3.11: External rectangle and neighboring background rectangle of a small target. Arrow V_T denotes the motion direction of the target. w represents target width while h stands for target height.

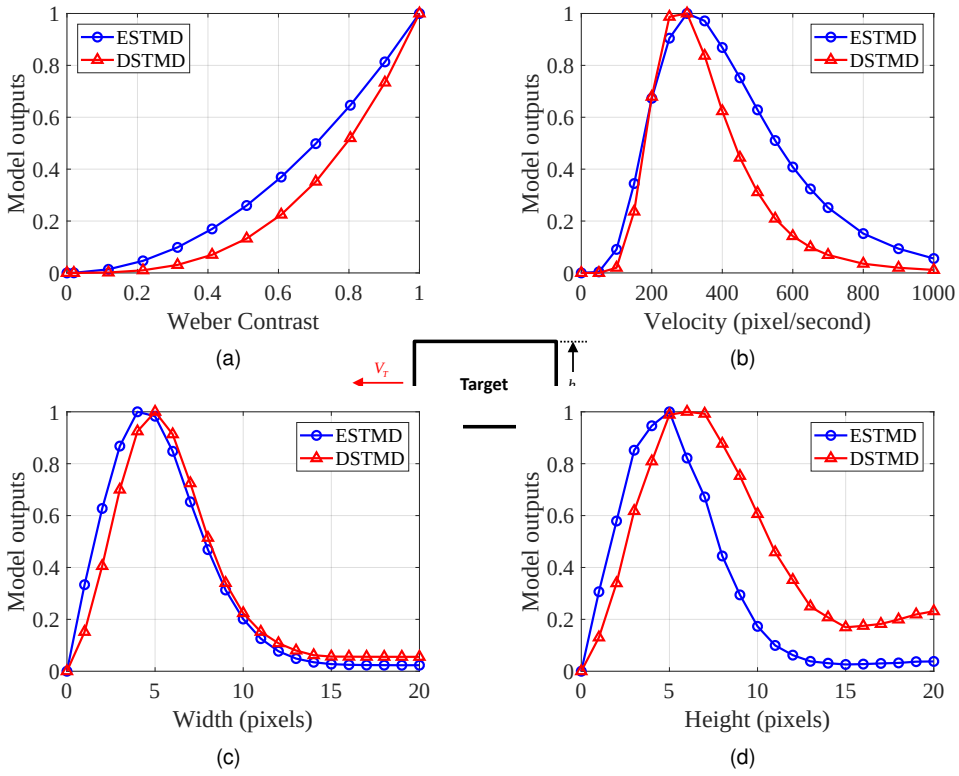


Figure 3.12: Tuning properties of the DSTMD and ESTMD. In each subplot, the horizontal axis represents one of target parameters (Weber Contrast, velocity, width and height) while the vertical axis denotes normalized model outputs. (a) Weber Contrast tuning curves. (b) Velocity tuning curves. (c) Width tuning curves. (d) Height tuning curves.

DSTMD exhibits the strongest output along direction $\theta = \pi$ which is consistent with the motion direction of the small target. The other seven outputs of the DSTMD decrease as the corresponding direction θ deviates from the small target motion direction. These eight directionally selective outputs are further used to encode motion direction of the small target by a population vector algorithm.

3.2.2 Tuning Properties

Four basic properties of the proposed neural model are tested in this section, including Weber Contrast sensitivity, velocity selectivity, width selectivity and height selectivity, which have been used to distinguish the STMD neurons in biology [13], [17], [29]. Here, Weber Contrast sensitivity refers to that the STMD neural response increases with the increase of Weber Contrast. Velocity selectivity refers to that the STMD neurons show the strongest response to a specific velocity (optimal velocity). Above or below this optimal velocity will result in the significant decrease of neural responses. Width selectivity and height selectivity are similar to velocity selectivity.

The definitions of Weber Contrast, width and height are given as follows. Width (or height) represents the target length extended parallel (or orthogonal) to the motion direction, as shown in Fig. 3.11. If the size of a target is $w \times h$, the size of its background rectangle is $(w + 2d) \times (h + 2d)$, where d is a constant which equals to 10 pixels. Then Weber Contrast is defined by the following equation,

$$\text{Weber Contrast} = \frac{|\mu_t - \mu_b|}{255} \tag{3.23}$$

where μ_t is the average pixel value of the target, μ_b is the average pixel value in neighboring area around the target.

We perform four experiments to illustrate four basic properties of the DSTMD and ESTMD. Image sequences which display a small target moving against the white background, are used as model input in the experiments. The initial parameters of the small target including luminance, velocity, width and height, are set as 0, 250 pixel/s, 5 pixels and 5 pixels, respectively. In each experiment, we change one of four target parameters while fix other three parameters, then record corresponding model outputs. The recorded tuning curves are displayed in Fig. 3.12.

As it is shown in Fig. 3.12(a), the outputs of the DSTMD and ESTMD increase with the increase of Weber Contrast, until reach maximum at Weber Contrast $= 1$. This reveals that the DSTMD and ESTMD exhibit Weber Contrast sensitivity. In Fig. 3.12(b), the outputs of the two models all peak at velocity $= 300$ pixel/s and decrease

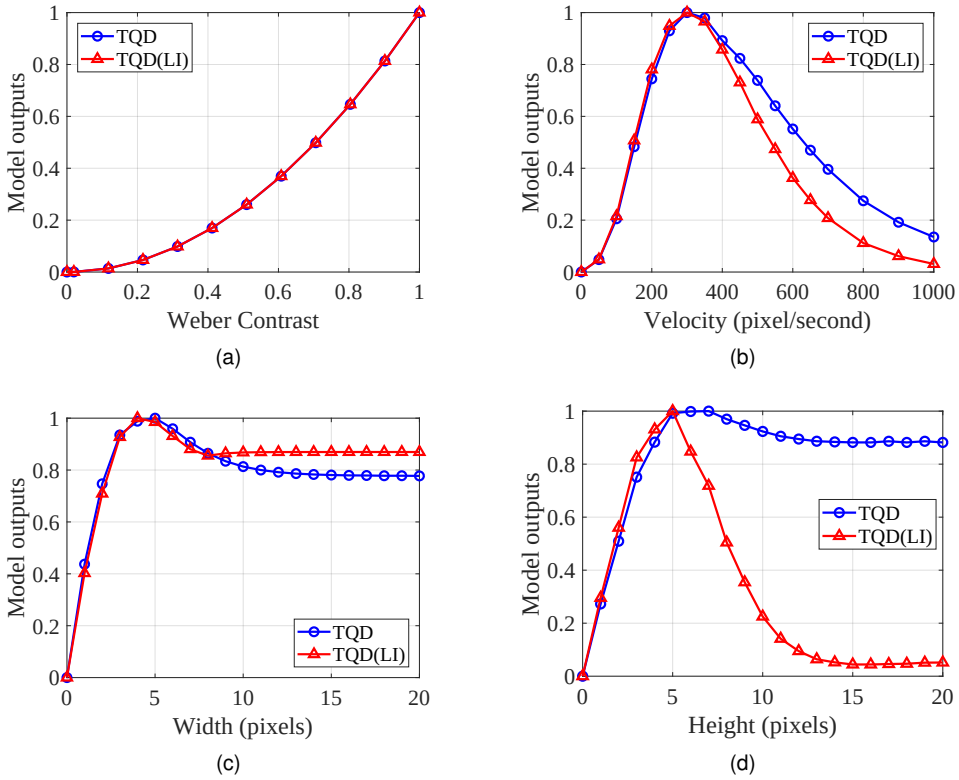


Figure 3.13: Tuning properties of the TQD and TQD(LI). In each subplot, the horizontal axis represents one of target parameters (Weber Contrast, velocity, width and height) while the vertical axis denotes normalized model outputs. (a) Weber Contrast tuning curves. (b) Velocity tuning curves. (c) Width tuning curves. (d) Height tuning curves.

significantly when the target velocity is above or below 300 pixel/s. This suggests that the DSTMD and ESTMD have a preferred velocity and exhibit velocity selectivity. Similar variation trends can be seen in Fig. 3.12(c) and (d) which reveal the width selectivity and height selectivity of the DSTMD, respectively.

We further carry out an experiment to demonstrate the hypothesis made in Section 3.1.3. The hypothesis is that if the medulla neurons which provide signals to the LPTC neurons [222]–[224], are laterally inhibited, the LPTC neurons would show strong size selectivity. In order to demonstrate this point, we first adopt TQD model [12], [93] to simulate the LPTC neurons. Then the medulla neuron modelling methods of DSTMD and ESTMD are used to simulate medulla neurons, respectively. For TQD which receives signals from medulla neurons modeled by DSTMD, we denote it as TQD. For TQD which receives signals from medulla neurons modeled by ESTMD, we denote it as TQD(LI). The only difference between the TQD and TQD(LI) is that medulla neurons providing signals to the TQD(LI), are laterally inhibited. Finally, the

four basic properties of the TQD and TQD(LI) are tested and their recorded tuning curves are presented in Fig. 3.13.

As can be seen from Fig. 3.13(a), (b) and (c), the TQD and TQD (LI) display minor differences. They all exhibit Weber Contrast sensitivity and velocity selectivity, but do not show width selectivity. In Fig. 3.13(d), although both TQD and TQD (LI) have a local maximum at height $= 5$, they show differences as the continuous increase of the height. Specifically, the output of the TQD firstly has a slight drop and finally tends to be stable around 0.9; in contrast, the output of the TQD (LI) decreases significantly and finally tends to be stable around 0.05. Above results indicate that the TQD(LI) exhibits height selectivity. This contradicts with the biological finding that the LPTC neurons are not size selective [222]–[224]. To avoid conflict with the biological finding on the LPTC neurons, we adopt the new medulla neuron modeling method and implement the lateral inhibition mechanism on the STMD neuron pathways.

3.2.3 Parameter Sensitivity

In the last section, we have demonstrated the four basic properties of the developed DSTMD, i.e., Weber Contrast sensitivity, velocity selectivity, width selectivity and height selectivity. In this section, we further evaluate the impacts of three sets of parameters on these four basic properties, including (n_4, τ_4), (n_5, τ_5) and (σ_4, σ_5) which are defined in (3.13) and (3.18). To assess the effects of these three sets of parameters, three experiments are conducted, each of which changes one set of parameters while keep other two sets of parameters at their initially assigned value (see Table 3.1). The tuning curves of the DSTMD under different parameter settings, are recorded and presented in Fig. 3.14, 3.15 and 3.16.

The first and second experiment reveal that the parameter (n_4, τ_4) and (n_5, τ_5) have large impact on the velocity selectivity and width selectivity, but show little effect on the Weber Contrast sensitivity and height selectivity. As it can be seen from Fig. 3.14(a) and (d), the increase of (n_4, τ_4) have not induced any significant changes of the Weber Contrast tuning curve and the height tuning curve. However, with the increase of (n_4, τ_4), the peak velocity decreases while the peak width increases, as displayed in

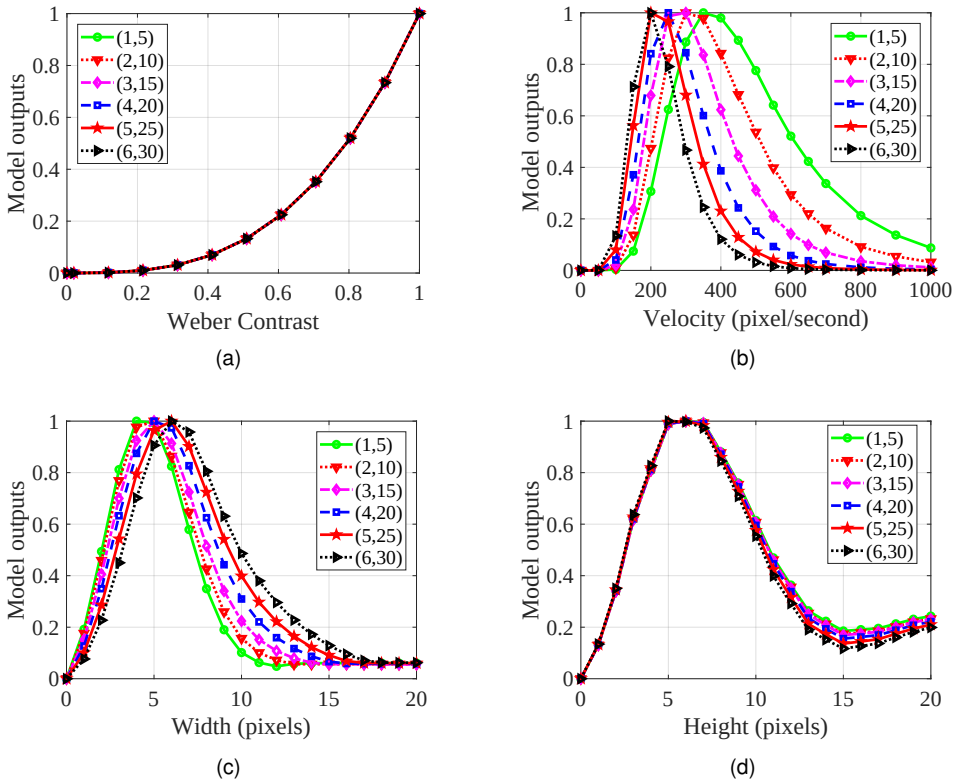


Figure 3.14: Tuning properties of the proposed neural model under different parameter (n_4, τ_4). In this experiment, (n_4, τ_4) is set as $(1, 5)$, $(2, 10)$, $(3, 15)$, $(4, 20)$, $(5, 25)$, $(6, 30)$ while the other parameters are fixed. In each subplot, the horizontal axis represents one of the target parameters (Weber Contrast, velocity, width and height) while the vertical axis denotes normalized model outputs. (a) Weber Contrast tuning curves. (b) Velocity tuning curves. (c) Width tuning curves. (d) Height tuning curves.

Fig. 3.14(b) and (c). In Fig. 3.15, the parameter (n_5, τ_5) has similar effect with (n_4, τ_4) on the four basic properties.

The reasons for the above results are - in the proposed neural model, τ_4 and τ_5 are positively correlated to $\frac{\alpha_1}{v}$ and $\frac{w}{v}$, respectively, where α_1, v and w stand for the distance between position A and B, the peak velocity and the peak width, respectively. Once α_1 is given, the increase of τ_4 (or τ_5) will result in the decrease of the peak velocity v and the increase of the peak width w.

We further explain why τ_4 and τ_5 are positively correlated to $\frac{\alpha_1}{v}$ and $\frac{w}{v}$. Fig. 3.17 presents the luminance changes of position A and B when a dark small target moves from B to A. In the equation (3.18), the DSTMD uses four medulla signals from position A (x, y) and B (x', y') to define the output of STMD neurons. Combining Fig. 3.17 with the equation (3.18), we point out that these four medulla signal are: 1) ON signal of position A (x, y), corresponding to $S^{\text{Tm3}}(x, y, t)$; 2) ON signal of position B (x', y')

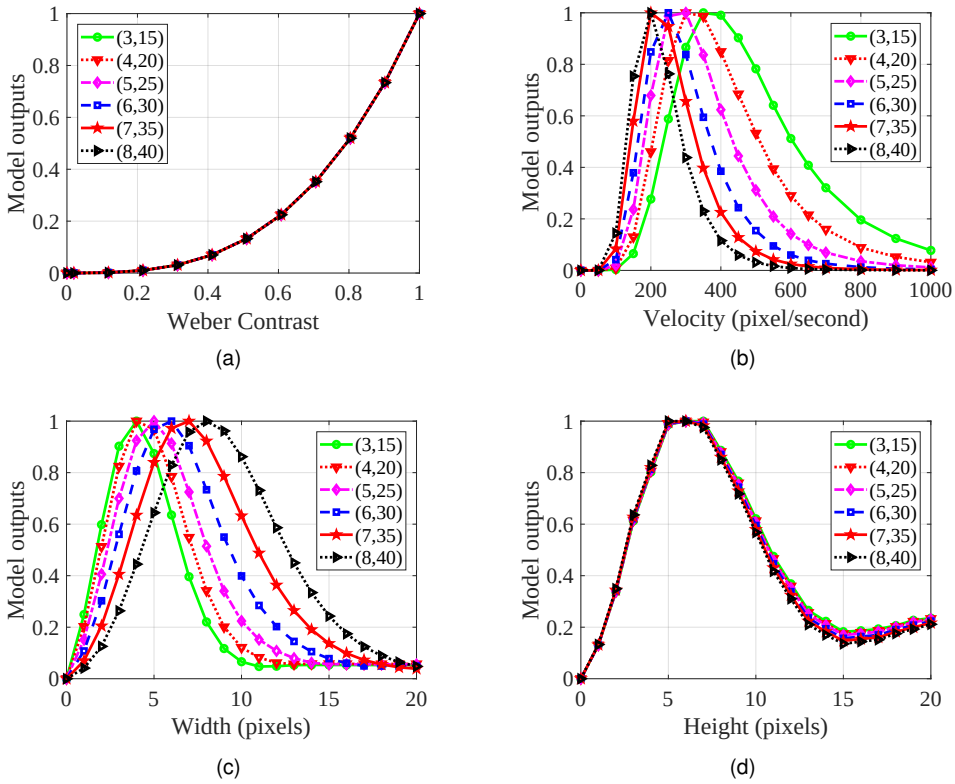


Figure 3.15: Tuning properties of the proposed neural model under different parameter (n_5, τ_5). In this experiment, (n_5, τ_5) is set as $(3, 15)$, $(4, 20)$, $(5, 25)$, $(6, 30)$, $(7, 35)$, $(8, 40)$ while the other parameters are fixed. In each subplot, the horizontal axis represents one of the target parameters (Weber Contrast, velocity, width and height) while the vertical axis denotes normalized model outputs. (a) Weber Contrast tuning curves. (b) Velocity tuning curves. (c) Width tuning curves. (d) Height tuning curves.

with time delay order n_4 and time delay length τ_4, corresponding to $S^{\text{Mi1}}_{(n_4,\tau_4)}(x', y', t)$; 3) OFF signal of position A (x, y) with time delay order n_5 and time delay length τ_5, corresponding to $S^{\text{Tm1}}_{(n_5,\tau_5)}(x, y, t)$; 4) OFF signal of position B (x', y') with time delay order n_6 and time delay length τ_6, corresponding to $S^{\text{Tm1}}_{(n_6,\tau_6)}(x', y', t)$. In the DSTMD, we set τ_4, τ_5 and τ_6 as Δt_1, Δt_2 and $\Delta t_1 + \Delta t_2$, respectively. Since $\Delta t_1 = \frac{\alpha_1}{v}$ and $\Delta t_2 = \frac{w}{v}$, then we have $\tau_4 = \frac{\alpha_1}{v}$ and $\tau_5 = \frac{w}{v}$. That is, τ_4 and τ_5 are positively correlated to $\frac{\alpha_1}{v}$ and $\frac{w}{v}$, respectively.

The third experiment demonstrates that the parameter (σ_4, σ_5) has large impact on the height selectivity, but shows little effect on the other three properties. In Fig. 3.16(a)-(c), the tuning curves have little changes with the increase of (σ_4, σ_5); in contrast, the peak height of the height tuning curve increases, as presented in Fig. 3.16(d). Here, we point out that the peak height is positively correlated to the size of the excitatory region of the lateral inhibition mechanism (see Fig. 3.6(b)). In the proposed

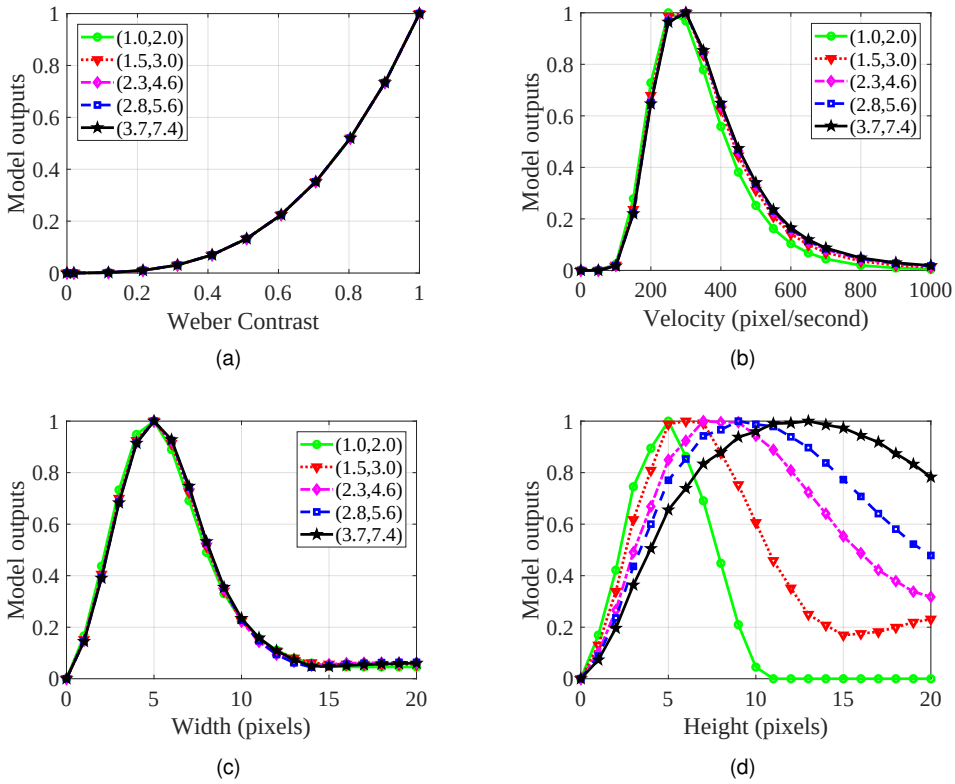


Figure 3.16: Tuning properties of the proposed neural model under different parameter (σ_4, σ_5). In this experiment, (σ_4, σ_5) is set as $(1.0, 2.0)$, $(1.5, 3.0)$, $(2.3, 4.6)$, $(2.8, 5.6)$, $(3.7, 7.4)$ while the other parameters are fixed. In each subplot, the horizontal axis represents one of the target parameters (Weber Contrast, velocity, width and height) while the vertical axis denotes normalized model outputs. (a) Weber Contrast tuning curves. (b) Velocity tuning curves. (c) Width tuning curves. (d) Height tuning curves.

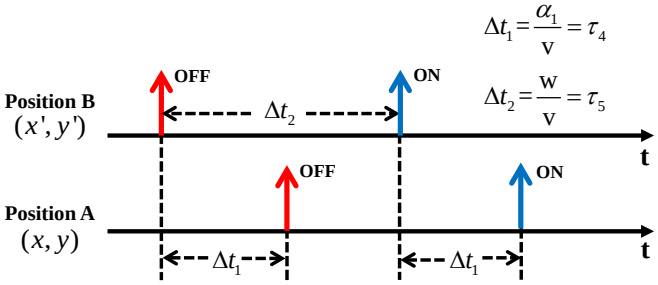
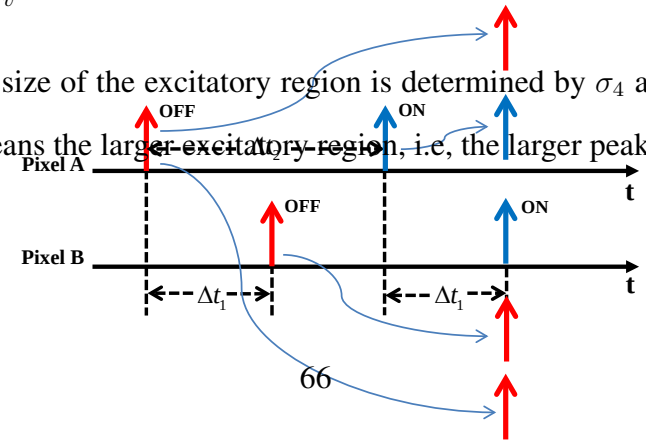


Figure 3.17: Schematic illustration of the luminance changes of the position A and B when a dark target successively passes position B (x', y') and A (x, y). The red arrow denotes luminance decrease signal (OFF signal) while the blue arrow represents luminance increase signal (ON signal). Let α_1, w and v stand for the distance between position A and B, target width and velocity, respectively. Then we have $\Delta t_1 = \frac{\alpha_1}{v}, \Delta t_2 = \frac{w}{v}$.

neural model, the size of the excitatory region is determined by σ_4 and σ_5, where the higher (σ_4, σ_5) means the larger excitatory region, i.e, the larger peak height.



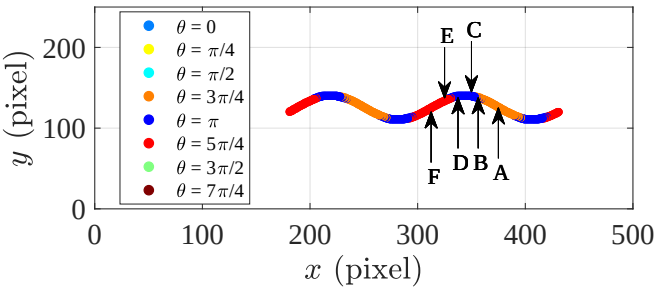


Figure 3.18: Motion trace of the small target where color denotes the direction of the strongest output of the proposed neural model.

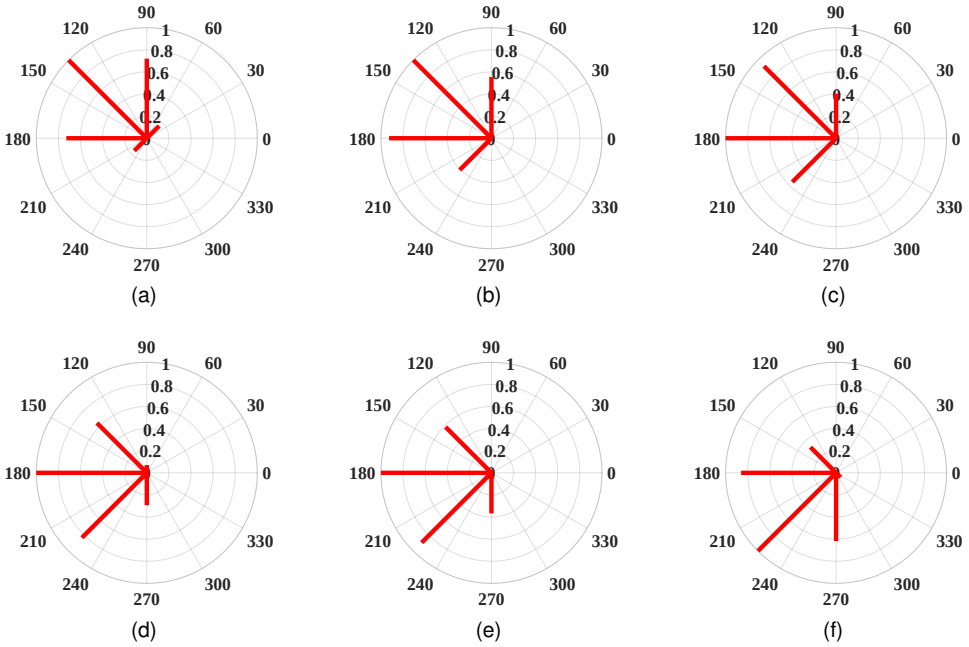


Figure 3.19: (a)-(f) Normalized DSTMD outputs at the position A,B,C,D,E,F. In each subplot, the angular coordinate represents the preferred motion direction of the DSTMD while the radial coordinate denotes the strength of the DSTMD output tuned to this preferred direction.

3.2.4 Direction Selectivity and Motion Direction Estimation

In this section, we illustrate how the proposed neural model encodes motion directions of small targets. An image sequence which displays a small target moving against the white background, is used as the model input. The luminance and size of the small target are set as 0 and 5×5 pixels, respectively. The coordinate of the small target at time t is $(500 - 250 \cdot \frac{t+300}{1000}, 125 + 15 \cdot \sin(4\pi \frac{t+300}{1000})), t \in [0, 1000]$ ms. Fig. 3.18 presents the motion trace of the small target. The motion direction of the small target varies between $142.98°$ and $217.01°$ when it moves along this motion trace.

We select six positions on the motion trace (A-F, in Fig. 3.18). The outputs of

Table 3.2: Estimated motion direction and actual motion direction at the six positions.

Position	Estimated Direction	Actual Direction	Difference
A	144.25°	143.12°	1.13°
B	152.36°	151.21°	1.15°
C	166.83°	166.88°	0.05°
D	180.37°	181.63°	1.26°
E	195.93°	197.80°	1.87°
F	214.24°	215.53°	1.29°

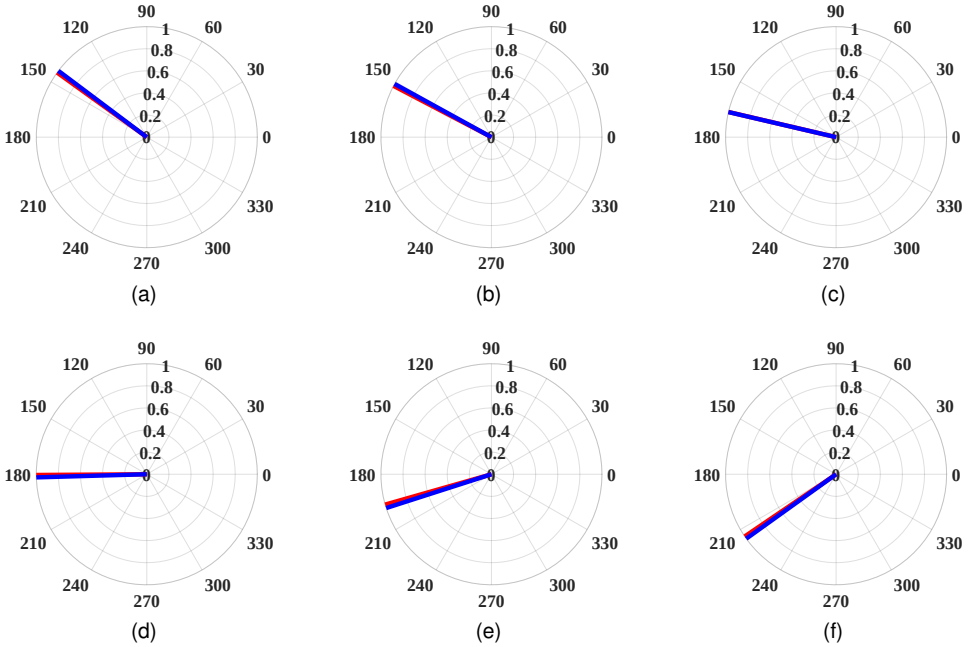


Figure 3.20: (a)-(f) Estimated motion direction (red) and actual motion direction (blue) at the position A,B,C,D,E,F. In each subplot, the red line is highly overlapped with the blue line. That is, the estimated motion direction is quite close to the actual motion direction.

the DSTMD at these six positions are normalized, then shown in polar coordinate (see Fig. 3.19). In each subplot of Fig. 3.19, we can see that the smaller difference between the preferred direction θ and actual motion direction (shown in Fig. 3.20), the stronger DSTMD output tuned to this direction θ. These directionally selective outputs are used to encode the motion direction of the small target by the population vector algorithm. Table 3.2 and Fig. 3.20 demonstrate the estimated motion direction and actual motion direction at the six positions. As can be seen, the difference between the estimated direction and actual direction is smaller than $2°$ at these six positions. We further estimate the motion direction of the small target at each position of the motion trace.

Table 3.3: Settings of the parameters including target luminance, size and horizontal velocity for the first three experiments.

	Luminance	Size	Velocity (V_T^x)
Experiment 1	$0, 25, 50$	5×5	250
Experiment 2	0	$3 \times 3, 5 \times 5, 8 \times 8$	250
Experiment 3	0	5×5	$200, 250, 350$

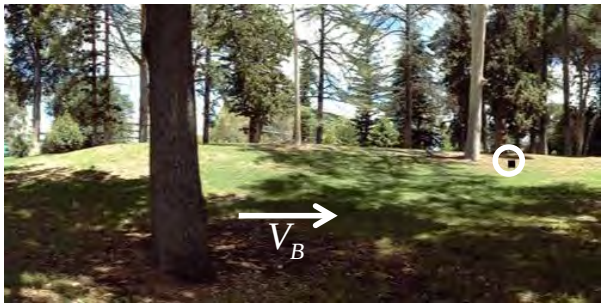


Figure 3.21: Representative frame of the input image sequence. The small target is highlighted by the white circle. The white arrow V_B denotes the motion direction of the background.

The maximal difference between the estimated motion direction and actual motion direction is $3.17°$. Above results indicate that the proposed neural model provides a good estimation for the motion direction of the small target.

3.2.5 Target Detection in Cluttered Backgrounds

The ability of the proposed neural model for detecting small targets against cluttered backgrounds is evaluated in this section. For a given detection threshold γ, if there is a position (x_0, y_0), time t_0 and direction θ_0 which satisfy the DSTMD output $E(x_0, y_0, t_0; \theta_0) > \gamma$, then we believe that a small target is detected at position (x_0, y_0) and time t_0. Two metrics are defined to evaluate the detection performance,

$$D_R = \frac{\text{number of true detections}}{\text{number of actual targets}} \tag{3.24}$$

$$F_A = \frac{\text{number of false detections}}{\text{number of images}} \tag{3.25}$$

where D_R and F_A denote the detection rate and false alarm rate, respectively. The detected result is considered correct if the pixel distance between the ground truth and the result is within a threshold (5 pixels).

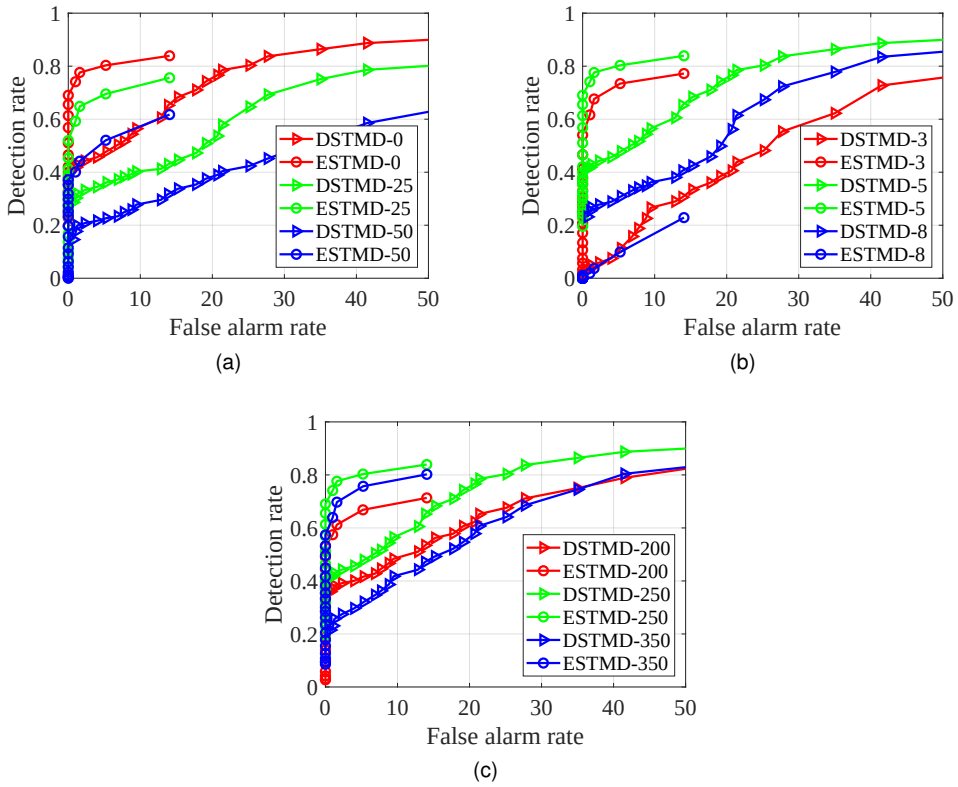


Figure 3.22: Receiver operating characteristic (ROC) curves of the first three experiments with respect to target luminance, sizes and velocities. (a) Experiment 1, different target luminance. Legend 'ESTMD-0' and 'DSTMD-0' represent the ROC curves of the ESTMD and DSTMD when target luminance equals to 0, respectively. Similar explanations for other legends. (b) Experiment 2, different target sizes. Legend 'ESTMD-3' and 'DSTMD-3' represent the ROC curves of the ESTMD and DSTMD when target size equals to 3×3 pixels, respectively. Similar explanations for other legends. (c) Experiment 3, different horizontal velocities (V_T^x). Legend 'ESTMD-200' and 'DSTMD-200' represent the ROC curves of the ESTMD and DSTMD when the horizontal velocity V_T^x equals to 200 pixel/s, respectively. Similar explanations for other legends.

In the first three experiments, we investigate the influences of three target parameters (size, luminance and velocity) on the detection performance. In each experiment, we change one of the target parameters while fix the other two parameters, then record the detection performance of the models under this parameter setting. The parameter settings of the first three experiments are shown in Table 3.3. All input image sequences are produced using the same background image where a representative frame is given in Fig. 3.21. In all input image sequences, the background is moving from left to right and its velocity is set as 250 pixel/s. A small target is moving against the cluttered background, and its coordinate at time t is $(500 - V_T^x \cdot \frac{t+300}{1000}, 125 + 15 \cdot \sin(4\pi\frac{t+300}{1000})), t \in [0, 1000]$ ms where V_T^x denotes the horizontal velocity. The receiver

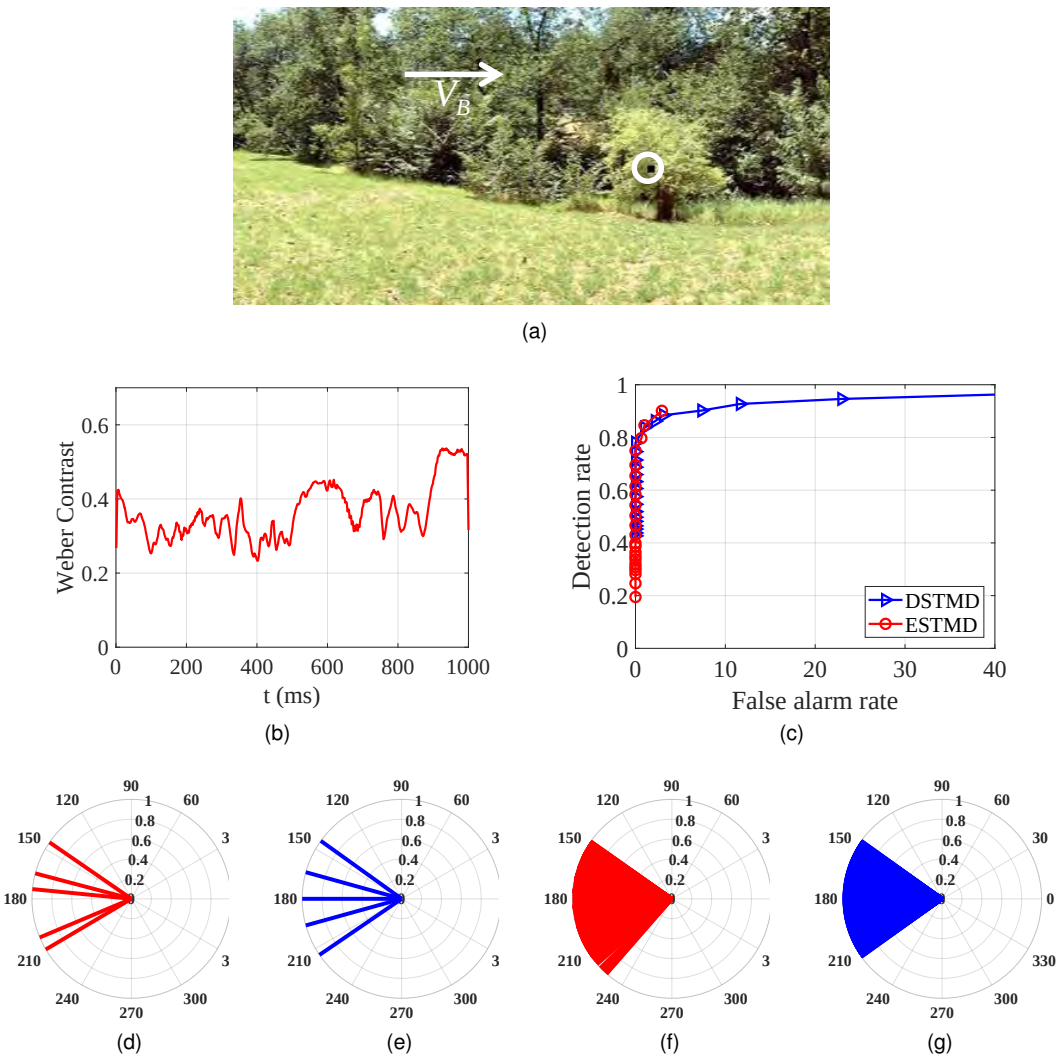


Figure 3.23: Experiment 4. (a) Representative frame of the input image sequence. (b) Weber Contrast of the small target during time period $t \in [0, 1000]$ ms. (c) ROC curves of the ESTMD and DSTMD. (d) Motion directions detected by the DSTMD in the sample 510, 570, 600, 630, 700 frames. No motion direction detected by the ESTMD. (e) Actual motion directions in the sample 510, 570, 600, 630, 700 frames. (f) Motion directions detected by the DSTMD from the 500th to the 700th frame. (g) Actual motion directions from the 500th to the 700th frame.

operating characteristic (ROC) curves of the three experiments with respect to target luminance, size and horizontal velocity V_T^x, are displayed in Fig. 3.22.

As can be seen from Fig. 3.22(a), the lower target luminance is, the better ESTMD and DSTMD perform. This is because the decrease of target luminance can induce the increase of Weber Contrast. Note that the ESTMD and DSTMD all show Weber Contrast sensitivity, so the higher Weber Contrast can elicit the stronger model output. From Fig. 3.22(b) and 3.22(c), we can see that when the false alarm rate is given, the target size of 5×5 (or the velocity of 250) has higher detection rate compared to

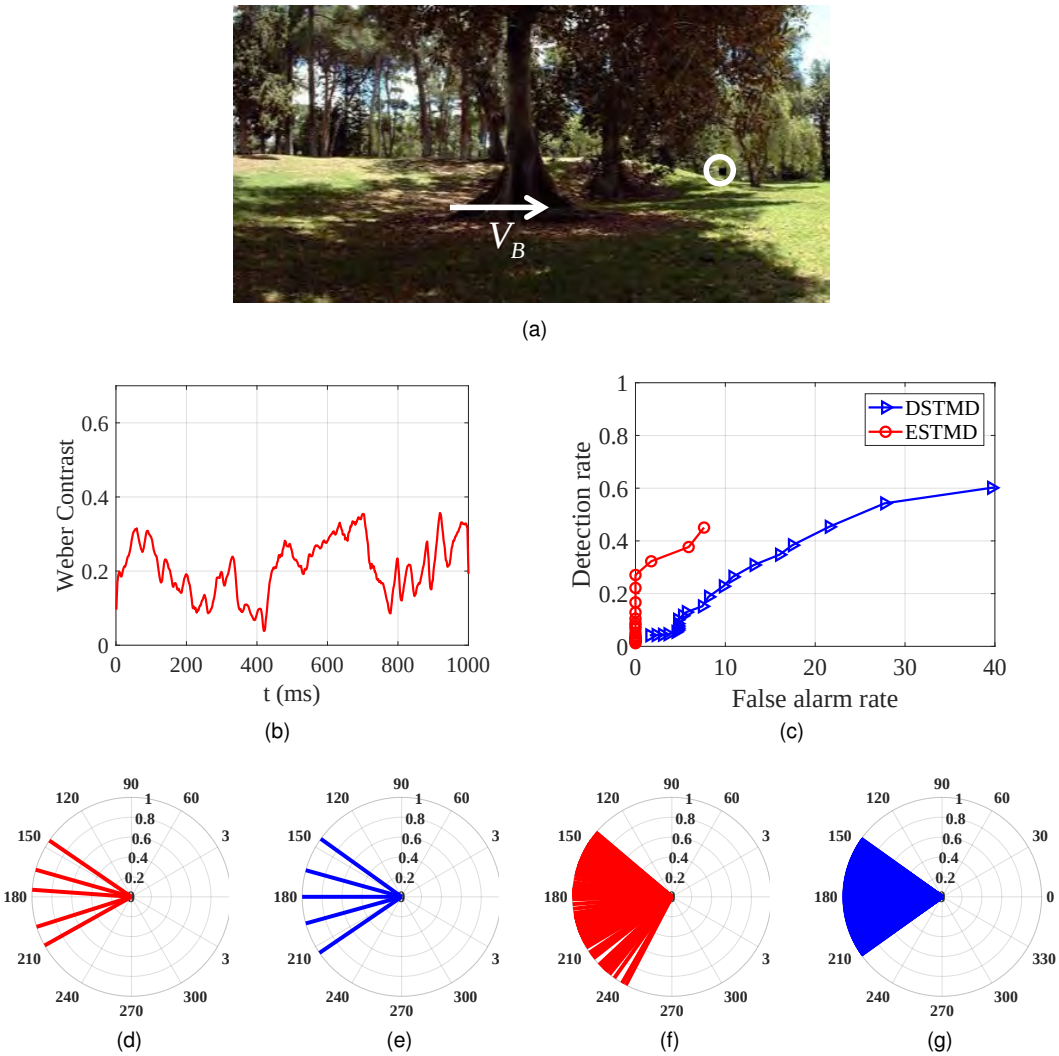


Figure 3.24: Experiment 5. (a) Representative frame of the input image sequence. (b) Weber Contrast of the small target during time period $t \in [0, 1000]$sms. (c) ROC curves of the ESTMD and DSTMD. (d) Motion directions detected by the DSTMD in the sample 510, 570, 600, 630, 700 frames. No motion direction detected by the ESTMD. (e) Actual motion directions in the sample 510, 570, 600, 630, 700 frames. (f) Motion directions detected by the DSTMD from the 500th to the 700th frame. (g) Actual motion directions from the 500th to the 700th frame.

the target size of 3×3 and 8×8 (or the velocity of 200 and 350). This is because the ESTMD and DSTMD all exhibit size and velocity selectivities. They show the strongest output to the target whose size (or velocity) equals to 5×5 pixels (or 250 pixel/s), but weaker outputs to the object whose size (or velocity) is higher or lower than 5×5 pixels (or 250 pixel/s). It can also be observed that the ESTMD performs better than the DSTMD in all the three experiments. For a given detection rate, the false alarm rate of the DSTMD is higher than that of the ESTMD. This is because the DSTMD has eight outputs at each pixel by correlating signals on eight directions

while the ESTMD only has a single output obtained by the correlation of signals from a single pixel. The correlation along eight directions can introduce direction selectivity, but it also brings additional false positives compared to the ESTMD.

In the fourth and fifth experiment, we evaluate the performance of the proposed neural model in different backgrounds. Two input image sequences with different backgrounds are displayed in Fig. 3.23(a) and Fig. 3.24(a), respectively. In these two image sequences, the backgrounds are all moving from left to right and their velocities are set as 250 pixel/second. A small target whose luminance and size are set as 0 and 5×5 pixels, is moving against the cluttered backgrounds. The coordinate of the small target at time t is $(500 - 250 \cdot \frac{t+300}{1000}, 125 + 15 \cdot \sin(4\pi \frac{t+300}{1000})), t \in [0, 1000]$ ms.

Fig. 3.23(c) and Fig. 3.24(c) demonstrate the ROC curves for the two image sequences, respectively. As can be seen, the detection rates of the DSTMD (or ESTMD) in Fig. 3.24(c) are much lower than those in Fig. 3.23(c). There are two reasons for the above result: 1) the background in Fig. 3.24(a) is more cluttered compared to Fig. 3.23(a), which means that it contains more small-target-like background features; 2) the Weber Contrast in Fig. 3.24(b) is much lower than that in Fig. 3.23(b), suggesting that the models exhibit much weaker outputs to the small target in the fifth experiment.

Fig. 3.23(d) displays the motion directions detected by the DSTMD in the sample 510, 570, 600, 630, 700 frames while Fig. 3.23(f) illustrates the motion directions detected by the DSTMD from the 500th to the 700th frame. As it is shown, these detected motion directions are quite close to the actual motion directions in Fig. 3.23(e) and 3.23(g). No motion direction is detected by the ESTMD, because it is not directionally selective. Similar results can be seen in Fig. 3.24(d)-(g).

3.2.6 Tests on Real Image Sequences

We further evaluate the performance of the proposed neural model on a set of real image sequences. Fig. 3.25 illustrates two representative frames of an real image sequence. In the experimental scene, the camera is stationary. A black ball is pulled by the experimenter with a transparent line in order to simulate small target motion. At the same time, pedestrians are walking against the background and regarded as noise.

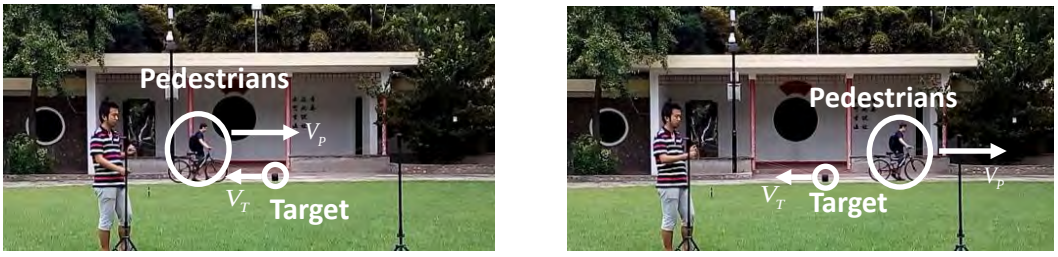


Figure 3.25: (a) The 805th frame of the recorded image sequence. (b) The 861th frame of the recorded image sequence. In the experimental scene, a small black ball which is highlighted in the white circle, is pulled by the experimenter with a transparent line. The pedestrians walking against the background are also highlighted in the white circle. The arrows V_T and V_P denote the motion directions of the small black ball and pedestrians, respectively.

Table 3.4: Size and velocity of the black ball and pedestrians in three real image sequences.

Real Image Sequence	1	2	3
Size of the black ball (pixels)	5×5	5×5	6×6
Size of the pedestrians (pixels)	15×20	25×50	20×20
Velocity of the black ball (pixel/frame)	0.16	0.21	0.25
Velocity of pedestrians (pixel/frame)	0.424	2.52	3.0

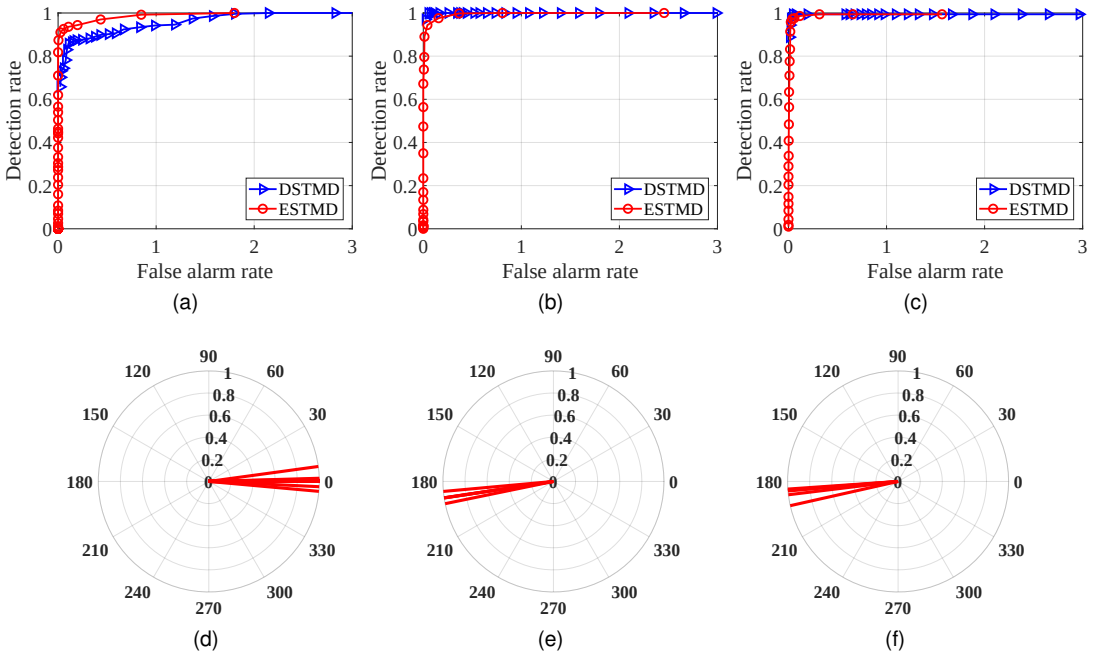


Figure 3.26: (a)-(c) ROC curves of the ESTMD and DSTMD for the three real image sequences. (d)-(f) Motion directions detected by the DSTMD in the sample 510, 570, 600, 630, 700 frames for the three real image sequences. No motion direction detected by the ESTMD.

The sizes and velocities of the black ball and the pedestrians in three real image sequences are listed in Table 3.4. Fig. 3.26 presents the ROC curves for the three real image sequences. It can be seen that both the DSTMD and ESTMD perform well on these three image sequences. More precisely, the DSTMD and ESTMD have a high detection rate at a low false alarm rate. This is because the sizes and velocities of the pedestrians in three recorded image sequences largely exceed the preferred ranges of the DSTMD and ESTMD. In this case, the DSTMD and ESTMD show much weaker responses to the pedestrians, but respond strongly to the black ball. Fig. 3.26(d)-(f) display the motion directions detected by the DSTMD in the sample 510, 570, 600, 630 and 700 frames. Since the black ball is pulled along a transparent line, its motion direction is close to 0 or π in these three image sequences. Due to the lack of direction selectivity, the ESTMD cannot detect the motion direction of the black ball.

3.3 Further Discussions

In the previous sections, the presented neural model (DSTMD) demonstrated a reliable ability to detect small targets and motion directions against complex backgrounds. Nowadays, for vision-based mobile robots, their visual sensors are becoming more reliable while computation ability is more powerful. These make it possible for mobile robots, such as unmanned aerial vehicle (UAV), equipped with the presented neural model to detect small moving targets in the distance in the real world.

In the insects' visual system, numerous neurons work together to extract different cues from the real world. For example, the LMCs extract motion information while the amacrine cells capture contrast information from input visual signals [226], [227]. Integrating these two types of information may contribute to the improvement of detection performance of the STMD neurons in cluttered backgrounds. In the future, the cooperation of these specialized neurons needs to be taken into consideration.

A number of bio-inspired neural networks based on firing-rate methods, spiking neural networks or convolutional neural networks [228]–[230], have been used for target detection, tracking and navigation. Although these neural networks perform well, they cannot distinguish small target motion from large object motion. Detecting

target motion is relatively easy, but distinguishing different target motion in terms of the targets' sizes is more challenging and difficult. For example, a naturally cluttered background always contains small targets such as insects, and large objects such as bushes, trees or rocks. Due to the camera motion, these large objects are moving with the background. In this case, the above-mentioned neural networks can detect both small and large object motion, but cannot distinguish them.

In engineering, small target motion detection can be performed by infrared detection methods [23]. However, these infrared methods always require significant temperature differences between objects of interest (such as rockets and jets) and the background. This largely limits their application, because such significant temperature difference is rare in the natural world. Different from the infrared methods, the presented neural model uses normal images as input and provides a vision-based method for small moving target detection.

3.4 Chapter Summary

In this chapter, we proposed a visual neural model (DSTMD) to simulate the directionally selective STMD neurons. Direction selectivity is obtained by correlating signals from two positions while size selectivity is introduced by a lateral inhibition mechanism. Motion directions of detected targets are estimated by the population vector algorithm. Systematic experiments showed that the presented STMD-based neural model can detect not only small moving targets, but also motion directions against complex backgrounds.

The developed DSTMD model provides an effective solution for small target detection and motion direction estimation, but its output always contains a number of false positives which result from miscorrelation of two signals and/or small-target-like background features. This is a common limitation for the existing STMD-based models with motion information only for small target detection. The following chapter will make effort to filter out the false positives by combining motion information with other visual cues.

Chapter 4

STMD Plus – a STMD with Contrast Pathway

A directionally selective STMD-based model (DSTMD) has been proposed in the last chapter to detect small moving targets and motion directions against cluttered backgrounds. To this end, motion information[1] is first extracted from input luminance signals, then further processed to enhance small target motion and eliminate background motion. However, the DSTMD cannot discriminate small moving targets from small-target-like background features (as shown in Fig. 4.1), which means that its detection results may contain a large number of false positives. This is because (1) small-target-like background features are embedded in the cluttered background such as bushes, trees and/or rocks, (2) they are moving with the whole background due to a free flying animal/camera. In this case, these small-target-like features (named as fake features) cannot be simply filtered out by existing STMD-based models with **motion information only** for small target motion detection. To address this problem, other visual information, such as directional contrast[2], should be combined with motion information for distinguishing small targets from fake features.

In the insects' visual systems, multiple visual cues are extracted by different specialized neural circuits [25], [71], [231]. Multiple neural circuits could be coordinated

[1] Motion information refers to luminance changes of a pixel with respect to time. From the view of mathematics, it is equivalent to temporal derivative of a pixel.
[2] Directional contrast denotes luminance changes of a pixel along different spatial directions. From the view of mathematics, it corresponds to directional derivatives of a pixel.

Figure 4.1: A small target is moving in the cluttered natural background which contains a number of small-target-like features (or called fake features). The small target and fake features all appear as small dim speckles whose sizes vary from one pixel to a few pixels, since they are far away from the animal/camera.

to discriminate small target motion. For example, in the lamina layer, large monopolar cells (LMCs) [7], [217] have been described as temporal band-pass filters which extract motion information from luminance signals [34], [37], [132]; and amacrine cells (AMCs) [232]–[234] linked to multi adjacent ommatidia with thin extending fibers, may constitute a contrast pathway with their downstream neurons to extract directional contrast from luminance signals. Although the contribution from the AMCs to STMD neural circuits in insects is unknown, it is clear that with directional contrast and motion information together, an artificial vision system could discriminate small moving targets from fake features robustly.

Inspired by the above biological findings, this chapter proposes a visual system model (**STMD+**) to detect small target motion in cluttered moving backgrounds, where small targets are successfully discriminated from fake features by combining motion information with directional contrast. The rest of this chapter is organized as follows. In Section 4.1, we introduce our proposed visual system model. Section 4.2 provides extensive performance evaluation as well as comparisons against the existing models. Finally, we conclude this chapter in Section 4.3.

4.1 Formulation of the System

In this section, we first illustrate the proposed visual system model schematically, then elaborate on its components in following subsections. The proposed visual system model is composed of four subsystems, including ommatidia, motion pathway, contrast pathway and mushroom body [235], [236], as illustrated in Fig. 4.2(a). The luminance

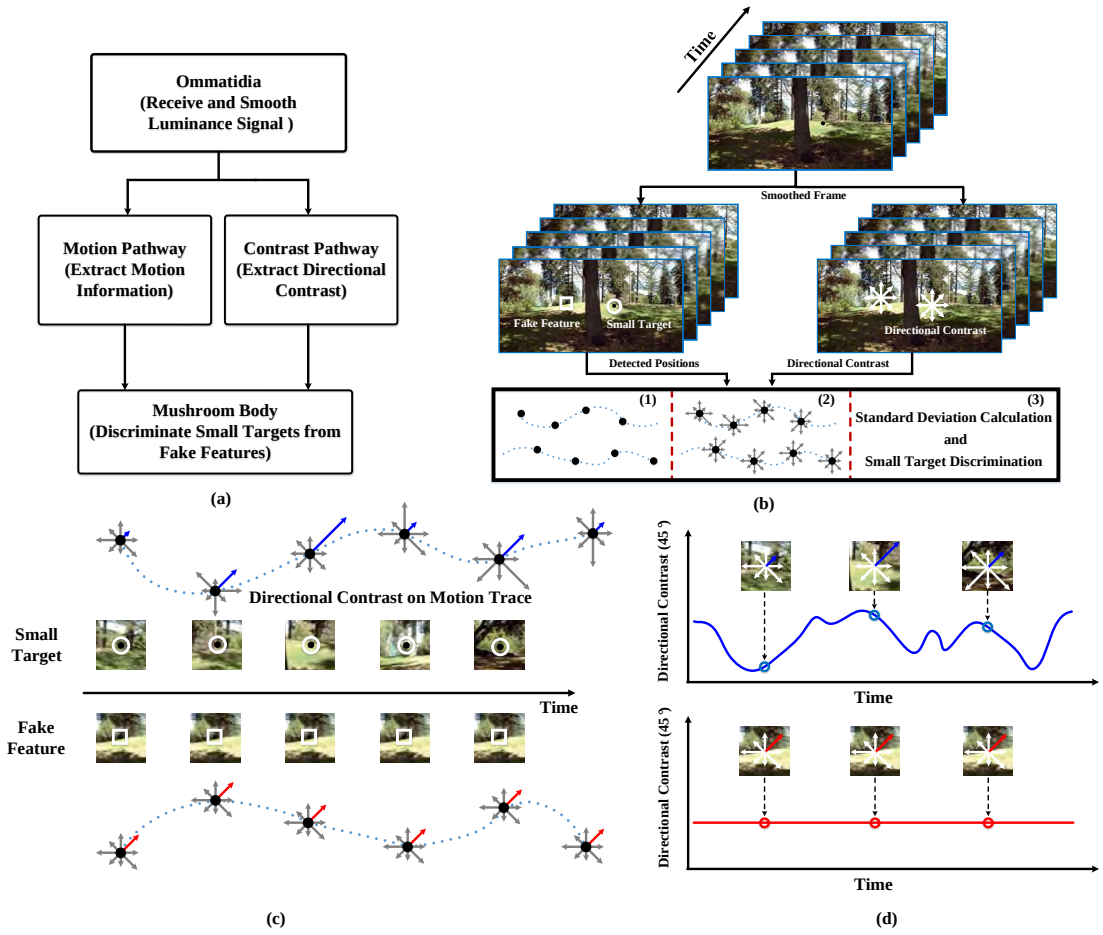
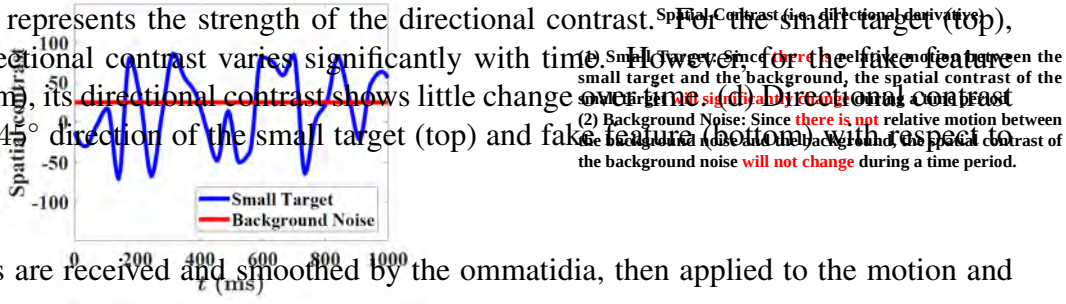


Figure 4.2: (a) Schematic illustration of the proposed visual system model (STMD+). (b) Image processing of the proposed visual system model. (c) Directional contrast on two motion traces which are caused by the small target and fake feature, respectively. Directional contrast is denoted by arrows along different directions where the arrow's length represents the strength of the directional contrast. For the small target (top), its directional contrast varies significantly with time. However, for the fake feature (bottom), its directional contrast shows little change over time. (d) Directional contrast along 45° direction of the small target (top) and fake feature (bottom) with respect to time.



signals are received and smoothed by the ommatidia, then applied to the motion and contrast pathways. These two pathways separately extract motion information and directional contrast which are finally integrated in the mushroom body to discriminate small targets from fake features.

Fig. 4.2(b) shows the image processing of the proposed visual system model, where the input image sequence is processed frame by frame. In each frame, both small targets and fake features are located by computing luminance changes of each pixel over time, while directional contrast is obtained by calculating luminance changes of each pixel along different directions. The detected positions and directional contrast

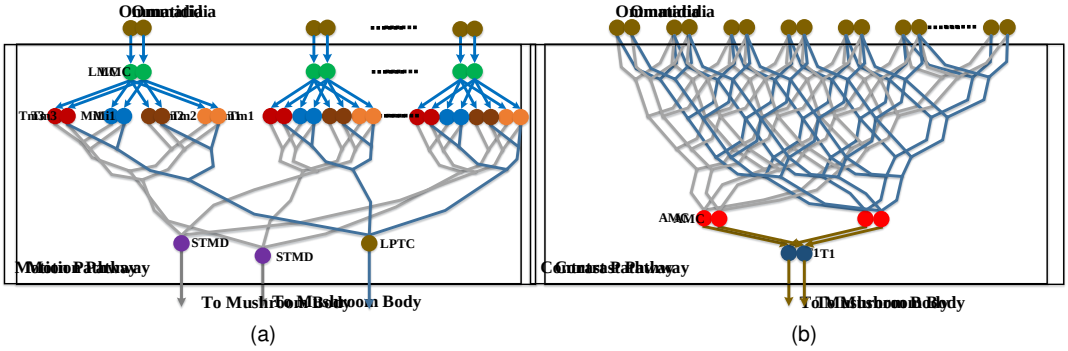
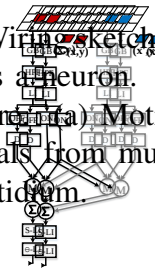


Figure 4.3: Wiring sketches of motion and contrast pathways. In subplots, each colored node denotes a neuron. For clear illustration, only one STMD and T1 neurons are presented here. (a) Motion pathway. (b) Contrast pathway. Note that each AMC collects signals from multiple ommatidia while each LMC receives signals from a single ommatidium.



are further processed as follows.

1. Successively record the detected positions to infer motion traces.

2. Extract the directional contrast on each motion trace.

3. Compute the standard deviation of directional contrast on each motion trace and compare it with a threshold for distinguishing small targets from fake features.

Our motivation is mainly based on the following observations: the directional contrast of small targets varies significantly with time, since they have relative movement to the background; on the contrary, the directional contrast of fake features shows little change over time, as they are static relative to the background. The variation amount in the directional contrast with time is represented by the standard deviation, which is taken as the criterion for small target discrimination. Fig. 4.2(c) visually displays the directional contrast on two typical motion traces that are separately caused by the small target and fake feature. As an example, Fig. 4.2(d) presents the directional contrast along $45°$ direction, which is used to calculate the standard deviation for this direction.

4.1.1 Ommatidia

Ommatidia act as luminance receptors to perceptive visual stimuli from the natural world [216]. In the proposed visual system, they are arranged in a matrix and modelled as spatial Gaussian filters, each of which captures and smooths the luminance of each

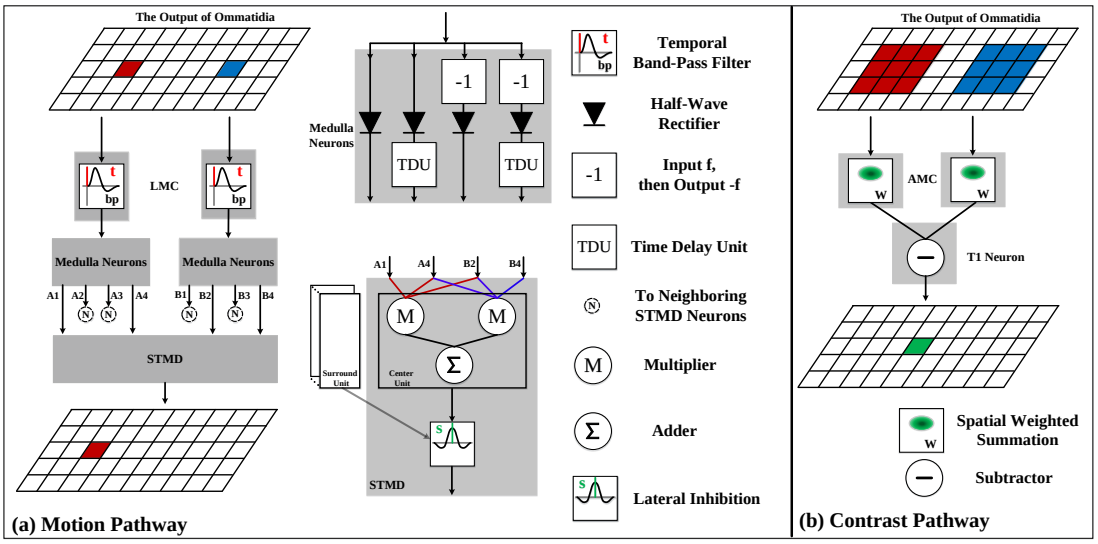


Figure 4.4: Schematic illustration of models of motion and contrast pathways. For clear illustration, only one STMD and T1 neurons are presented here. However, these types of neurons are all arranged in matrix form in the proposed visual system.

pixel in the input image. Formally, let $I(x, y, t) \in \mathbb{R}$ denote the input image sequence, where x, y and t are spatial and temporal field positions. The output of an ommatidium $P(x, y, t)$ is given by,

$$P(x, y, t) = \iint I(u, v, t) G_{\sigma_1}(x - u, y - v) du dv \tag{4.1}$$

where $G_{\sigma_1}(x, y)$ is a Gaussian function, defined as

$$G_{\sigma_1}(x, y) = \frac{1}{2\pi\sigma_1^2} \exp(-\frac{x^2 + y^2}{2\sigma_1^2}). \tag{4.2}$$

4.1.2 Motion Pathway

As shown in Fig. 4.3(a), the motion pathway consists of large monopolar cells (LMCs) [7], [217], medulla neurons (i.e., Mi1, Tm1, Tm2 and Tm3) [237], [238], small target motion detectors (STMDs) [17], [28], [239] and lobula plate tangential cells (LPTCs) [57], [222]. The output of ommatidia is first fed into LMCs, then processed by medulla neurons and finally integrated by STMDs. Fig. 4.4(a) displays the model of the motion pathway, which is elaborated as follows.

***1) Large Monopolar Cells (LMCs)*:** Objects' motion can induce luminance changes of pixels with time. These luminance changes are extracted by the LMCs, each of

which is modelled by a temporal band-pass filter that is defined as the difference of two Gamma kernels (see Fig. 4.4(a)). That is,

$$H(t) = \Gamma_{n_1,\tau_1}(t) - \Gamma_{n_2,\tau_2}(t) \tag{4.3}$$

$$\Gamma_{n,\tau}(t) = (nt)^n \frac{\exp(-nt/\tau)}{(n-1)! \cdot \tau^{n+1}} \tag{4.4}$$

where $H(t)$ denotes the impulse response of the band-pass filter, $\Gamma_{n,\tau}(t)$ stands for the Gamma kernel [218], n and τ refers to the order and time constant of the Gamma kernel $\Gamma_{n,\tau}(t)$. Then the output of each LMC can be calculated by convolving $H(t)$ with the output of ommatidia $P(x, y, t)$,

$$L(x, y, t) = \int P(x, y, s)H(t - s)ds. \tag{4.5}$$

The $L(x, y, t)$ reflects luminance changes of pixel (x, y) over time t, where a positive $L(x, y, t)$ means luminance increase while a negative $L(x, y, t)$ suggests luminance decrease.

***2) Medulla Neurons*:** Medulla neurons including Tm1, Tm2, Tm3 and Mi1, constitute four parallel channels to process the output of LMCs $L(x, y, t)$. The Tm3 and Tm2 are modelled as half-wave rectifiers to separate $L(x, y, t)$ into luminance increase and decrease components. Let $S^{\text{Tm3}}(x, y, t)$ and $S^{\text{Tm2}}(x, y, t)$ denote the output of the Tm3 and Tm2, respectively, then they are given by

$$S^{\text{Tm3}}(x, y, t) = [L(x, y, t)]^+ \tag{4.6}$$

$$S^{\text{Tm2}}(x, y, t) = [-L(x, y, t)]^+ \tag{4.7}$$

where $[x]^+$ denotes $\max(x, 0)$. The Mi1 and Tm1 further temporally delay $S^{\text{Tm3}}(x, y, t)$ and $S^{\text{Tm2}}(x, y, t)$ by convolving them with a Gamma kernel. That is,

$$S^{\text{Mi1}}_{(n,\tau)}(x, y, t) = \int [L(x, y, s)]^+ \cdot \Gamma_{n,\tau}(t - s)ds \tag{4.8}$$

$$S^{\text{Tm1}}_{(n,\tau)}(x, y, t) = \int [-L(x, y, s)]^+ \cdot \Gamma_{n,\tau}(t - s)ds \tag{4.9}$$

where $S^{\text{Mi1}}_{(n,\tau)}(x,y,t)$ and $S^{\text{Tm1}}_{(n,\tau)}(x,y,t)$ represent the outputs of the Mi1 and Tm1, respectively; n and τ are the order and time constant of the Gamma kernel, which separately determine the order and time-delay length of the time delay unit (TDU) (see Fig. 4.4(a)).

***3) Small Target Motion Detectors (STMDs)*:** As can be seen from Fig. 4.4(a), each STMD collects the outputs of medulla neurons located at two pixels, i.e., (x,y) and $(x'(\theta),y'(\theta))$ which are defined as

$$\begin{aligned} x'(\theta) &= x + \alpha_1 \cos\theta \\ y'(\theta) &= y + \alpha_1 \sin\theta \end{aligned} \tag{4.10}$$

where α_1 is a constant, θ denotes the preferred direction of the STMD. When a dim object successively moves over pixels (x,y) and $(x'(\theta),y'(\theta))$, a luminance decrease followed by a luminance increase will appear at each of these two pixels. These luminance increase and decrease signals are first aligned in time domain and then multiplied together so as to produce a large response [132]. That is,

$$\begin{aligned} D(x,y,t,\theta) = S^{\text{Tm3}}(x,y,t) \cdot \Big\{ & S^{\text{Tm1}}_{(n_4,\tau_4)}(x,y,t) \\ & + S^{\text{Mi1}}_{(n_3,\tau_3)}(x'(\theta),y'(\theta),t) \Big\} \cdot S^{\text{Tm1}}_{(n_5,\tau_5)}(x'(\theta),y'(\theta),t) \end{aligned} \tag{4.11}$$

where $D(x,y,t,\theta)$ denotes the output of the STMD neuron with a preferred direction θ. Here, θ belongs to $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}\}$, corresponding to eight preferred directions of STMD neurons (see Fig. 4.5). It is worthy to note that τ_3, τ_4 and τ_5 are determined by the different delays among the luminance changes, while n_3, n_4 and n_5 are accordingly tuned to guarantee appropriate Gamma kernel shapes [132].

So far, the obtained $D(x,y,t,\theta)$ can detect both small and large moving objects in the forms of producing a large response. In order to suppress the responses to large moving objects, the $D(x,y,t,\theta)$ is further laterally inhibited by convolving with an inhibition kernel $W_s(x,y)$. That is,

$$E(x,y,t,\theta) = \iint D(u,v,t,\theta) W_s(x-u, y-v) du dv \tag{4.12}$$

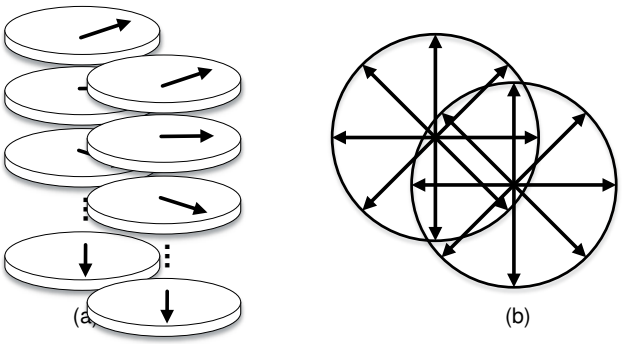


Figure 4.5: (a) Illustration of neurons which are located at the same position, but have different preferred directions. The black arrows denote preferred directions. (b) Illustration of different preferred directions in the x-y plane.

where $E(x, y, t, \theta)$ represents the inhibited signal; the inhibition kernel $W_s(x, y)$ is defined as

$$W_s(x, y) = A \cdot [g(x, y)]^+ + B \cdot [g(x, y)]^- \tag{4.13}$$

$$g(x, y) = G_{\sigma_2}(x, y) - e \cdot G_{\sigma_3}(x, y) - \rho \tag{4.14}$$

where $[x]^+$ and $[x]^-$ respectively denote $\max(x, 0)$ and $\min(x, 0)$; A, B, e and ρ are constant.

By comparing the $E(x, y, t, \theta)$ with a detection threshold β, we can find the positions of small moving objects. Specially, if $E(x, y, t, \theta) > \beta$, then we believe that a small object moving along direction θ is located at pixel (x, y) and time t. However, it cannot distinguish small targets and fake features that can be both recognized as small moving objects. To address this issue, we construct a contrast pathway accounting for directional contrast calculation.

4.1.3 Contrast Pathway

As shown in Fig. 4.3(b), contrast pathway is composed of amacrine cells (AMCs) [232]–[234] and T1 neurons [240], [241]. The output of ommatidia is firstly fed into AMCs, then processed by T1 neurons. Fig. 4.4(b) displays the model of the contrast pathway, which is elaborated as follows.

***1) Amacrine Cells (AMCs)*:** Each AMC receives the output of multiple ommatidia located in a small region and serves as a weighted summation unit, as presented in Fig.

4.4(b). Here, we define the weight function as

$$W_{_A}(x,y) = \frac{1}{2\pi\eta^2}\exp(-\frac{x^2+y^2}{2\eta^2}) \tag{4.15}$$

where η is constant. Then the output of each AMC $A(x,y,t)$ can be given by

$$A(x,y,t) = \iint P(u,v,t)W_{_A}(x-u,y-v)dudv \tag{4.16}$$

where $P(x,y,t)$ is the output of ommatidia defined in (4.1).

***2) T1 Neurons*:** The T1 neuron layer is adopted to extract the directional contrast along different directions. The directional contrast at (x,y) along direction ϕ is defined as the difference between the outputs of two AMCs that are located at $(x+\alpha_2\cos\phi, y+\alpha_2\sin\phi)$ and $(x-\alpha_2\cos\phi, y-\alpha_2\sin\phi)$. Here, α_2 is a constant. Let $T(x,y,t,\phi)$ denote the output of a T1 neuron with a preferred direction ϕ, then it can be given by

$$\begin{aligned} T(x,y,t,\phi) =& A(x+\alpha_2\cos\phi, y+\alpha_2\sin\phi, t) \\ &- A(x-\alpha_2\cos\phi, y-\alpha_2\sin\phi, t). \end{aligned} \tag{4.17}$$

Substituting (4.16) in (4.17), we have

$$T(x,y,t,\phi) = \iint P(u,v,t)W_{_T}(x-u,y-v,\phi)dudv \tag{4.18}$$

where the convolution kernel $W_{_T}(x,y,\phi)$ represents

$$\begin{aligned} W_{_T}(x,y,\phi) =& W_{_A}(x+\alpha_2\cos\phi, y+\alpha_2\sin\phi) \\ &- W_{_A}(x-\alpha_2\cos\phi, y-\alpha_2\sin\phi). \end{aligned} \tag{4.19}$$

Here ϕ belongs to $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$, corresponding to four preferred directions of T1 neurons. It is worthy to note that the convolution kernel $W_{_T}(x,y,\phi)$ is one of the directional derivative operators [242], [243], which can extract anisotropic luminance variations (see Fig. 4.6).

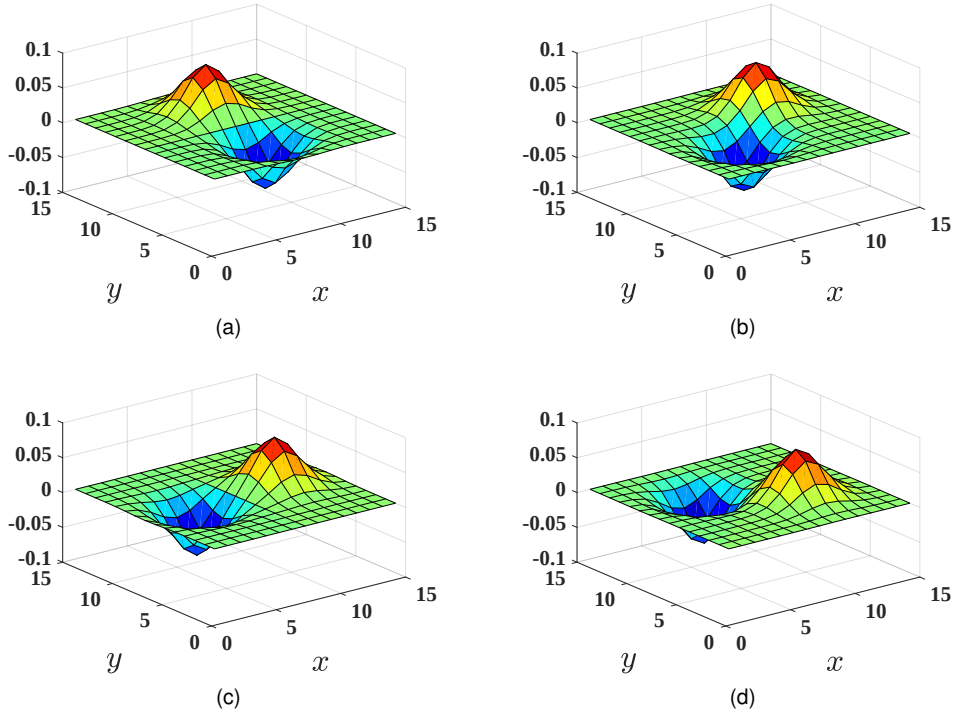


Figure 4.6: Illustration of the convolution kernel $W_{_T}(x, y, \phi)$. (a) $\phi = 0$. (b) $\phi = \frac{\pi}{4}$. (c) $\phi = \frac{\pi}{2}$. (d) $\phi = \frac{3\pi}{4}$.

4.1.4 Mushroom Body

In the proposed visual system, the mushroom body [235], [236] receives two types of neural outputs, including the output of STMDs $E(x, y, t, \theta)$ and the output of T1 neurons $T(x, y, t, \phi)$. These neural outputs are integrated to discriminate small targets from fake features via the following three procedures.

***1) Motion Trace Recording*:** The output of STMDs $E(x, y, t, \theta)$ is employed to record motion traces of small objects. For a detection threshold β and a starting time t_0, if there exists a pixel A $(x_{_A}, y_{_A})$ and motion direction $\theta_{_A}$ which satisfy $E(x_{_A}, y_{_A}, t_0, \theta_{_A}) > \beta$, then we believe that a small object[3] is detected at pixel $(x_{_A}, y_{_A})$ and its motion direction is $\theta_{_A}$. Similarly, at next time step t_1, another pixel B $(x_{_B}, y_{_B})$ and motion direction $\theta_{_B}$ can be detected. Especially, if the pixel B $(x_{_B}, y_{_B})$ is located in the small neighborhood of the pixel A $(x_{_A}, y_{_A})$, and pixel B is the nearest detected point to pixel A, then we believe that pixels A and B belong to the same motion trace denoted by TR. Repeating the above steps, the motion trace TR can be recorded

[3]The detected object could be a small target or a fake feature, which cannot be discriminated by the STMDs.

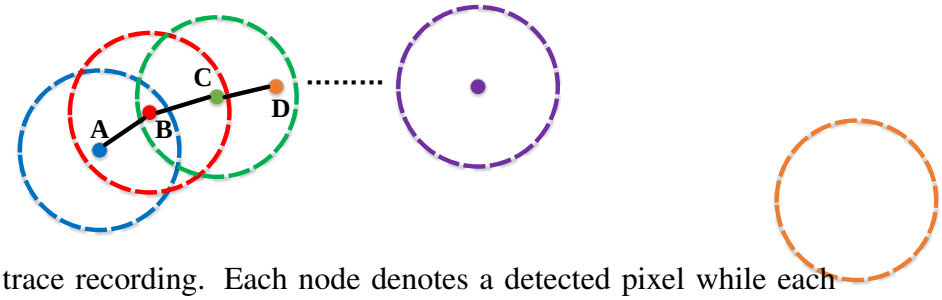


Figure 4.7: Motion trace recording. Each node denotes a detected pixel while each circle represents a small neighborhood. If the pixel B is located in the neighborhood of the pixel A, and pixel B is the nearest detected point to pixel A, then we believe that pixels A and B belong to the same motion trace. Repeating this step, a motion trace could be recorded.

during a time period, as shown in Fig. 4.7. The TR can be described as,

$$TR = (x(t), y(t), \theta(t)), t \in [t_0, t_n] \tag{4.20}$$

where $x(t)$ and $y(t)$ represent x and y coordinates at time t, $\theta(t)$ denotes motion direction, t_0 and t_n are the starting time and current time.

***2) Information Integration*:** Once motion traces are recorded, we can obtain their directional contrast by substituting (4.20) into $T(x, y, t, \phi)$. That is,

$$Q(t, \phi) = T(x(t), y(t), t, \phi), t \in [t_0, t_n] \tag{4.21}$$

where $Q(t, \phi)$ denotes the directional contrast along direction ϕ on the motion trace TR; $(x(t), y(t))$ stands for the point on the motion trace. To quantify the variation amount in the directional contrast, we calculate the standard deviation (SD) of the $Q(t, \phi)$ during a time period $[t_{n-m}, t_n]$, denoted by $SD(t_{n-m}, t_n, \phi)$. Here m represents the sample number for the SD calculation.

***3) Small Target Discrimination*:** We determine whether a detected object is a small target or a fake feature, using the standard deviations of the directional contrast on the object's motion trace, i.e., $SD(t_{n-m}, t_n, \phi)$. If the $SD(t_{n-m}, t_n, \phi)$ is smaller than a certain threshold, we believe that the detected object is a fake feature; Otherwise, it is a small target.

Algorithm 1 Detection Process of the STMD Plus

Input: Image sequence $\{I_1, I_2, \cdots, I_N\}$, where $I_i \in \mathbb{R}^{m \times n}$.
Output: Positions of small moving targets in each input image.

for each input image **do**
  // Ommatidia
  Calculate the output of the ommatidium via (4.1);
  // Motion Pathway
  Calculate the output of the LMC via (4.5);
  Calculate the outputs of the medulla neurons via (4.6)-(4.9);
  Calculate the output of the STMD via (4.11);
  Calculate the laterally inhibited output via (4.12);
  // Contrast Pathway
  Calculate the output of the AMC via (4.16);
  Calculate the output of the T1 neuron via (4.17);
  // Mushroom Body
  Calculate motion traces of the detected objects via (4.20);
  for each motion trace **do**
    Calculate the directional contrast of the motion trace via (4.21);
    Calculate the standard deviations (SD) of the directional contrast on the motion trace;
    if $SD >$ threshold **then**
      the detected object is a small target;
    else
      the detected object is a fake feature.
    end if
  end for
end for

4.1.5 Computational Complexity

For demonstration of actual implementations, we attach pseudo-code form of the STMD Plus (see Algorithm 1). We further briefly discuss the complexity of the proposed method for small target motion detection. As shown in Algorithm 1, the computational time of our method mainly consists of four parts: the ommatidia, the motion pathway, the contrast pathway and the mushroom body.

The computational complexity of the ommatidia is essentially determined by a 2-D spatial convolution of the input image with a Gaussian kernel (see Equation (4.1)), which can be implemented in $O(k^2mn)$ time for an $m \times n$ input image and a $k \times k$ kernel.

In the motion pathway, the LMC output can be regarded as the difference of two Gamma convolutions (see Equation (4.3)-(4.5)). Since the temporal Gamma convolution needs $O(lmn)$ cost where l is the length of the Gamma kernel, the computa-

tional complexity of the LMC scales with $O(2lmn)$. Similarly, the total cost of the four medulla neurons is about $O(2lmn + 2mn)$. According to (4.11), the computational complexity of the STMD is $O(2mn)$ for each preferred direction, so its entire cost grows like $O(2dmn)$ where d denotes the number of the preferred directions. Finally, the lateral inhibition mechanism which is implemented by a 2-D convolution (see Equation (4.12)), needs $O(k^2mn)$ cost. Thus the entire computational complexity of the motion pathway is $O((k^2 + 4l + 2d + 2)mn)$.

In the contrast pathway, the directional contrast of each pixel along different spatial directions is calculated by convolving the ommatidium output with directional derivative operators (see Equation (4.17)). Since the 2-D spatial convolution needs $O(k^2mn)$ cost for each spatial direction, the entire computational complexity of the contrast pathway is $O(k^2dmn)$.

In the mushroom body, the nearest neighbour of each detected object is calculated for recording motion trace, which can be implemented in $O(p^2)$ time in the worst case where p is the number of the detected objects. In addition, the cost of standard deviation calculation is around $O(rp)$ where r represents the sample number. So the entire computational complexity of the mushroom body is around $O(p^2 + rp)$.

Based on the aforementioned analysis, the entire computational complexity of the proposed STMD Plus is around $O(N(2k^2 + k^2d + 4l + 2d + 2)mn + N(p^2 + rp))$ where N stands for the number of input images. As can be seen, the front part of the computational complexity will be fixed once the kernel size k, l and the number of directions d are given. However, the latter part grows as the square of the p, which is very computationally intensive for a large p. Note that the cluttered background always contains a number of false positives, so the p could be quite large when a low detection threshold is chosen. Due to the high computational complexity in motion trace recording, the STMD Plus may not be suitable for real-time applications.

4.1.6 Parameter Setting

Parameters of the proposed visual system model are listed in Table 4.1, where the parameters of the motion pathway are determined by the analysis in [132] while those

Table 4.1: Parameters of the proposed visual system model.

Eq.	Parameters
(4.1)	$\sigma_1 = 1$
(4.3)	$n_1 = 2, \tau_1 = 3, n_2 = 6, \tau_2 = 9$
(4.10)	$\alpha_1 = 3$
(4.11)	$n_3 = 3, \tau_3 = 15, n_4 = 5, \tau_4 = 25, n_5 = 8, \tau_5 = 40$
(4.13)	$A = 1, B = 3$
(4.14)	$\sigma_2 = 1.5, \sigma_3 = 3.0, e = 1, \rho = 0$
(4.15)	$\eta = 1.5$
(4.17)	$\alpha_2 = 3$

of the contrast pathway are tuned based on empirical experience. These parameters are chosen to satisfy the functionality, which are mainly determined by the velocity and size ranges of the moving targets. They will not be changed in the following experiments unless stated.

The proposed visual system model is written in Matlab (The MathWorks, Inc., Natick, MA). The computer used in the experiments is a standard laptop with a 2.50GHz Intel Core i7 CPU and 16GB DDR3 memory. The source code can be found at https://github.com/wanghongxin/STMD-Plus.

4.2 Results and Discussions

The proposed visual system model is evaluated on a synthetic dataset [225] and a real dataset (STNS dataset) [37]. The synthetic dataset contains a number of image sequences which are synthesized by using real background images and a computer generated small target (a black block). These image sequences all display the motion of the small target against the cluttered moving backgrounds, which are different in the target sizes, target velocities, background velocities, background types and so on. The sampling frequencies of the synthetic videos are all equal to 1000 Hz. The STNS dataset is a collection of 25 real videos featuring various moving targets and environments. The scenarios include many kinds of challenges, such as heavy clutter, camera motion and changes in overall brightness. The STNS dataset (videos and manual ground truth annotations) is available at https://figshare.com/articles/STNS_Dataset/4496768.

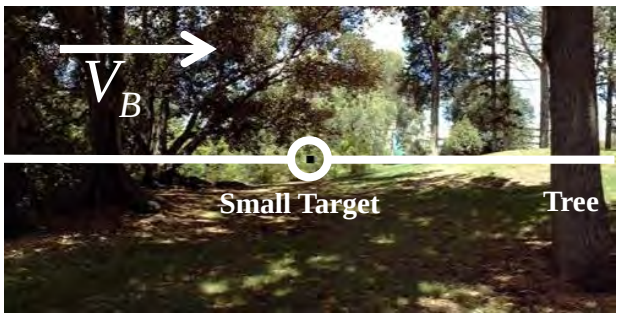


Figure 4.8: A input frame at time $t_0 = 1000$ ms whose resolution is 500 pixels (in horizontal) by 250 pixels (in vertical). The small target (the black block) and the cluttered background are moving from left to right. Their velocities are all equal to 250 pixel/s, where arrow V_B denotes the motion direction of the background. The tree which is regarded as a large object, is also moving due to the background motion.

To quantitatively evaluate the detection performance, two metrics are defined as following [23],

$$D_R = \frac{\text{number of true detections}}{\text{number of actual targets}} \tag{4.22}$$

$$F_A = \frac{\text{number of false detections}}{\text{number of images}} \tag{4.23}$$

where D_R and F_A denote detection rate and false alarm rate, respectively. The detected result is considered correct if the pixel distance between the ground truth and the result is within a threshold (5 pixels).

4.2.1 Signal Processing in the Motion Pathway

To intuitively illustrate the signal processing in the motion pathway, we observe the output of each neural layer with respect to x by setting y and t as $y_0 = 125$ pixel and $t_0 = 1000$ ms. Fig. 4.8 shows the input frame at time $t_0 = 1000$ ms, where the luminance signal $I(x, y_0, t_0)$ on the middle line is presented in Fig. 4.9(a). Its resulting ommatidium output and LMC output are displayed in Fig. 4.9(b) and (c), respectively. The ommatidum output is a smoothed version of the input signal. The LMC output reveals the luminance changes of pixels, where the positive values correspond to luminance increase while the negative values suggest luminance decrease.

Fig. 4.10(a) demonstrates the four inputs of the STMDs when the preferred direction θ is set to 0. Specifically, the $S^{\text{Tm3}}(x, y_0, t_0)$ is the positive part of the LMC

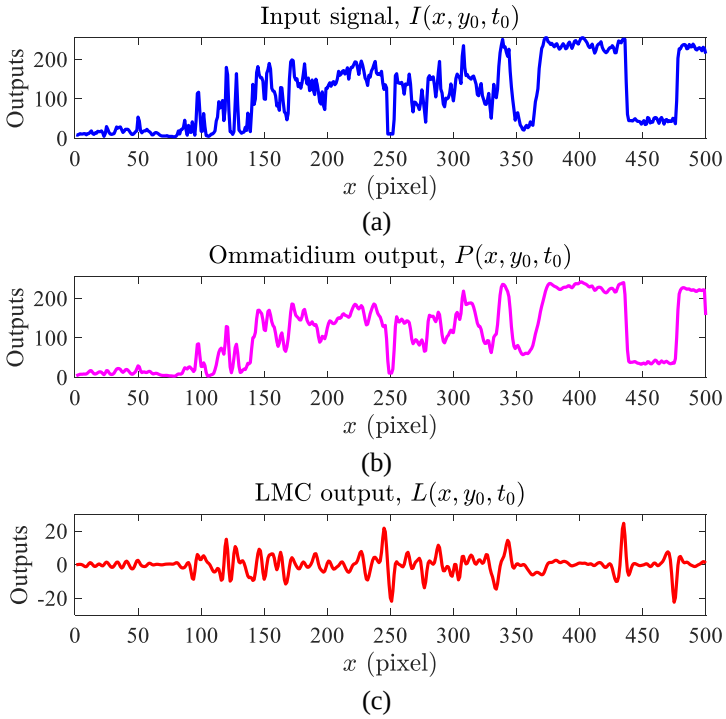
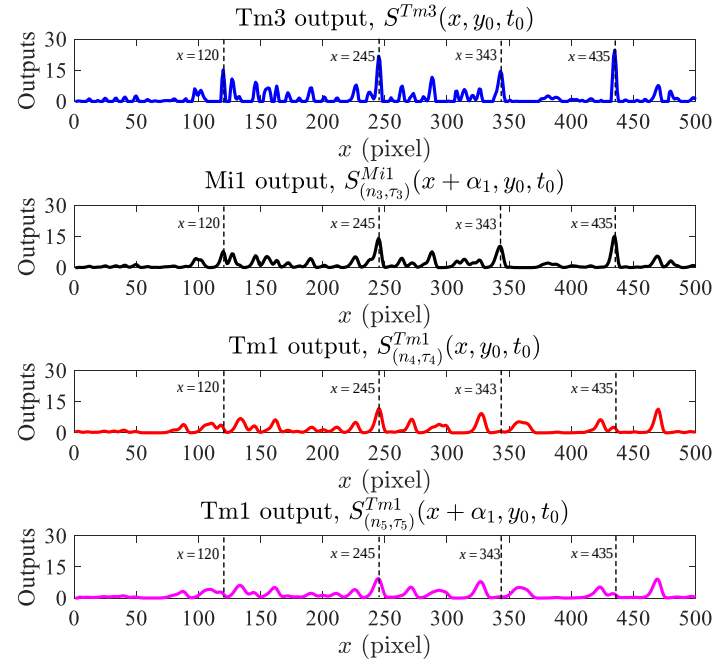


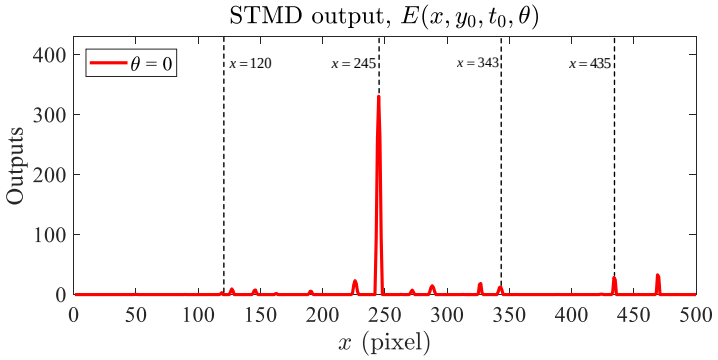
Figure 4.9: In each subplot, the horizontal axis denotes x coordinate while the vertical axis represents neural outputs. (a) Input luminance signal $I(x, y_0, t_0)$. (b) Ommatidium output $P(x, y_0, t_0)$. (c) LMC output $L(x, y_0, t_0)$.

output; the $S^{\text{Mi1}}_{(n_3,\tau_3)}(x + \alpha_1, y_0, t_0)$ denotes the delayed version of the positive part of the LMC output with a shift of α_1 pixels; the $S^{\text{Tm1}}_{(n_4,\tau_4)}(x, y_0, t_0)$ stands for the delayed version of the negative part of the LMC output; and the $S^{\text{Tm1}}_{(n_5,\tau_5)}(x + \alpha_1, y_0, t_0)$ represents the delayed version of negative part of the LMC output with a shift of α_1 pixels. Fig. 4.10(b) further shows the output of STMDs, where a high response appears at the position of the small target ($x = 245$) while the responses at other positions are effectively suppressed. This is because the four peaks located at the position of the small target are aligned (see Fig. 4.10(a)), which will produce a strong response after the multiplication, summation and lateral inhibition in the STMD (see Fig. 4.4). For other positions e.g., $x = 120, 343, 435$, the peaks on the four curves exhibit a low aligning probability, hence producing a weak response. Note that the lateral inhibition is introduced to suppress the responses to large objects, such as the tree displayed in Fig. 4.8.

It is worthy to note that the above analysis is based on the presetting of the preferred direction $\theta = 0$. When we change the preferred direction θ, different STMD outputs can be calculated. Fig. 4.11 presents the STMD outputs at the positions $x = 245$ and



(a)



(b)

Figure 4.10: In each subplot, the horizontal axis denotes x coordinate while the vertical axis represents neural outputs. (a) Four inputs of the STMDs when the preferred direction θ is set to 0, i.e., $S^{Tm3}(x, y_0, t_0)$, $S^{Mi1}_{(n_3,\tau_3)}(x+\alpha_1, y_0, t_0)$, $S^{Tm1}_{(n_4,\tau_4)}(x, y_0, t_0)$ and $S^{Tm1}_{(n_5,\tau_5)}(x+\alpha_1, y_0, t_0)$. (b) STMD output $E(x, y_0, t_0, \theta)$ when the preferred direction θ is equal to 0.

$x = 435$ along eight preferred directions θ, where $x = 245$ is the position of the small target and $x = 435$ corresponds to the position of the large tree. As shown in Fig. 4.11(a), for the small target, the STMD shows strong directional selectivity. As the preferred direction deviates from the motion direction of the small target, the STMD output will decrease correspondingly. On the other hand, the direction of the small target can be estimated by computing the summation of these output vectors [132]. For the large tree (see Fig. 4.11(b)), the outputs of the STMD along eight preferred directions are very low, suggesting that the STMD is not interested in large moving objects.

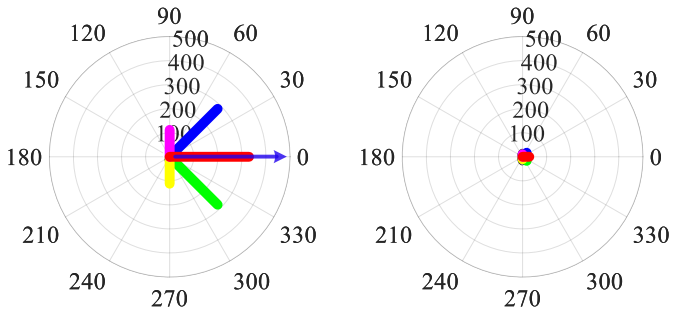


Figure 4.11: In the polar coordinate system, the angular coordinate represents the preferred direction θ while the radial coordinate denotes the STMD output. (a) STMD outputs at position $x = 245$ along eight preferred directions $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}\}$. The blue arrow stands for the motion direction of the small target. (b) STMD outputs at position $x = 435$ along eight preferred directions θ.

4.2.2 Characteristics of the STMD

To further demonstrate the characteristics of the STMD, we compare its outputs to objects with different velocities, widths, heights and Weber contrast. Fig. 4.12(a) shows the STMD output with respect to the Weber contrast. As can be seen, the STMD output increases as the increase of Weber contrast, until reaches maximum at Weber contrast $= 1$. This indicates that the higher Weber contrast of an object is, the easier it can be detected. Fig. 4.12(b) presents the STMD output with regard to the velocity of the moving object. Obviously, the STMD output peaks at an optimal velocity (300 pixel/s). The STMD also exhibits high responses to the objects whose velocities range from 100 to 500 pixel/s. Fig. 4.12(c) and (d) display the output of the STMD when changing the width and height of the object, which indicate that the STMD prefers moving objects whose widths and heights are smaller than 15 pixels.

These characteristics of the STMD revealed in Fig. 4.12(a)-(d), are called Weber contrast sensitivity, velocity selectivity, width selectivity and height selectivity, respectively, which have been already found in the STMD neurons in biological research [13], [17], [239].

4.2.3 Effectiveness of the Contrast Pathway

In the proposed visual system model, we design a contrast pathway and incorporate it with the motion pathway to discriminate small targets from fake features. To validate

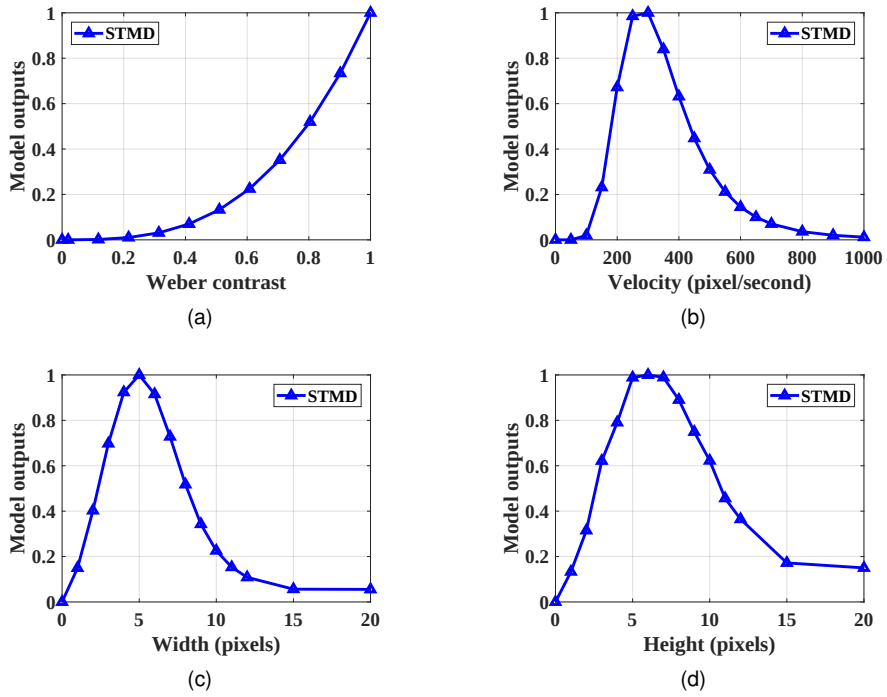


Figure 4.12: STMD outputs to moving objects with different Weber contrast, velocities, widths and heights. (a) Diffe
Different widths. (d) Different heigł

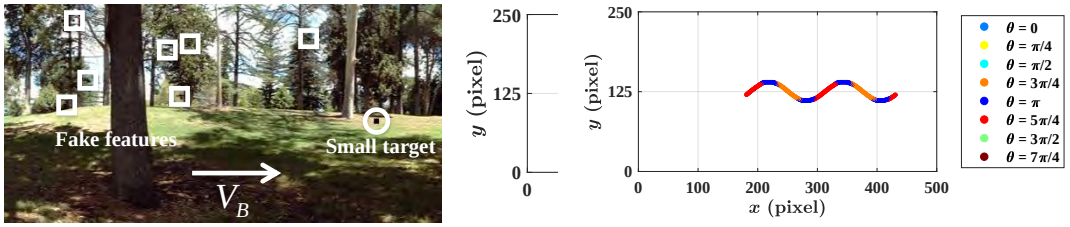


Figure 4.13: (a) Representative frame of the input image sequence. A small target (the small black block) highlighted by the circle, is moving against the cluttered background. The background which contains a number of fake features, is also moving from left to right where arrow V_B denotes the background motion direction. (b) Motion trace of the small target during time period $[0, 1000]$ ms, i.e., ground truth. In this subplot, color represents motion direction θ of the small target.

its effectiveness, we first compare the performance of the STMD+ with and without the contrast pathway. Then we conduct a performance comparison between the developed STMD+ and two baseline models including ESTMD [34] and DSTMD [132]. The testing setups are detailed as follows: the input image sequence is presented in Fig. 4.13(a), which displays a small target moving against the cluttered background; the background is moving from left to right and its velocity is 250 pixel/s; the luminance, size and velocity of small target are equal to 0, 5×5 pixels and 250 pixel/s, respectively;

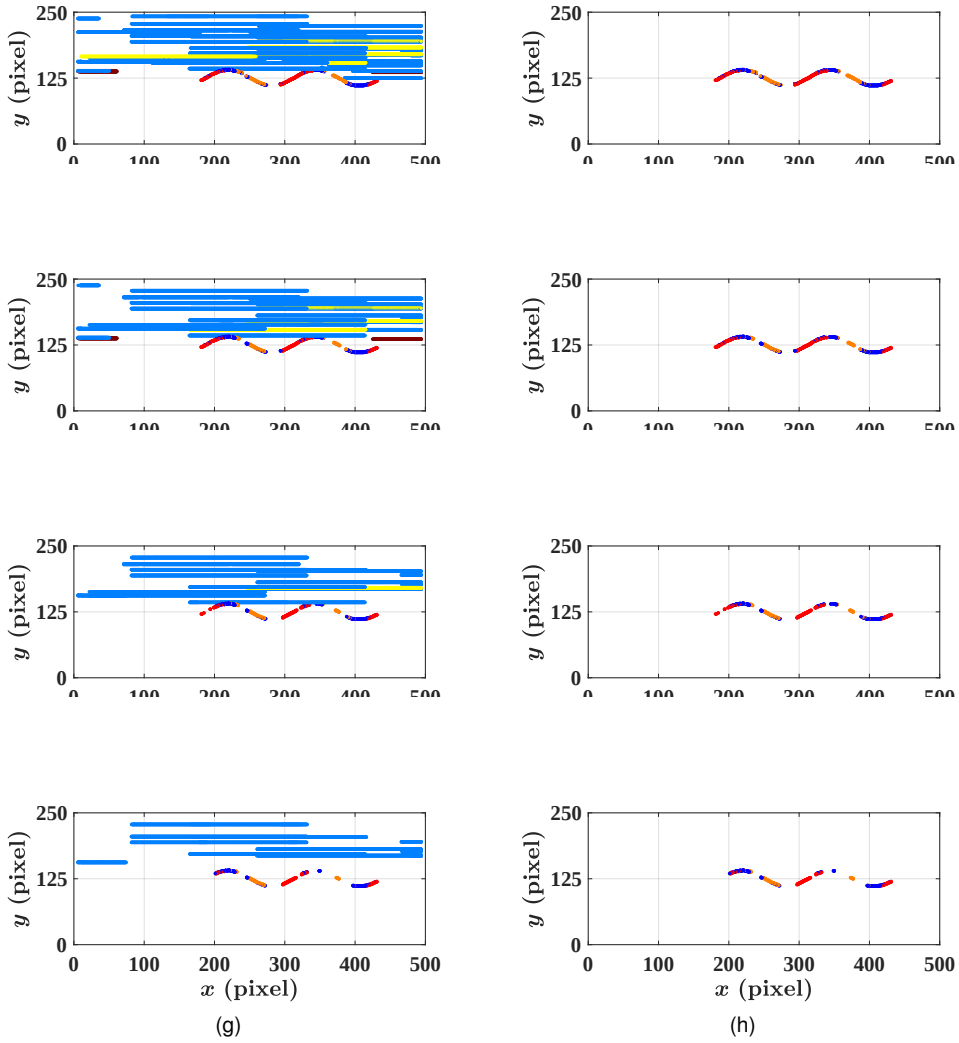


Figure 4.14: (a), (c), (e) and (g) Motion traces detected by the STMD+ without the contrast pathway under different detection thresholds β which are set as 150, 250, 350 and 450, respectively. (b), (d), (f) and (h) Motion traces detected by the STMD+ with the contrast pathway under different detection thresholds β which are set as 150, 250, 350 and 450, respectively.

the motion trace of the small target during time period $[0, 1000]$ ms is illustrated in Fig. 4.13(b).

Fig. 4.14(a), (c), (e) and (g) displays the motion traces detected by the STMD+ without the contrast pathway under different detection thresholds β. As can be seen, these detection results all contain numerous fake features. When increasing the detection threshold, the detected fake features will decrease while the detected motion trace becomes more incomplete. After applying the contrast pathway, the fake features are all filtered out even under different detection thresholds (see Fig. 4.14(b), (d), (f) and (h)). The specific detection rate (D_R) and false alarm rate (F_A) are presented Table 4.2.

Table 4.2: Detection rate (D_R) and false alarm rate (F_A) of the STMD+ with and without the contrast pathway under different detection thresholds β.

Threshold β	Without*		With#	
	D_R	F_A	D_R	F_A
150	**0.85**	**27.70**	**0.85**	**0**
250	0.74	19.11	0.74	0
350	0.60	12.87	0.60	0
450	0.50	6.88	0.50	0

* The STMD+ without the contrast pathway
the con

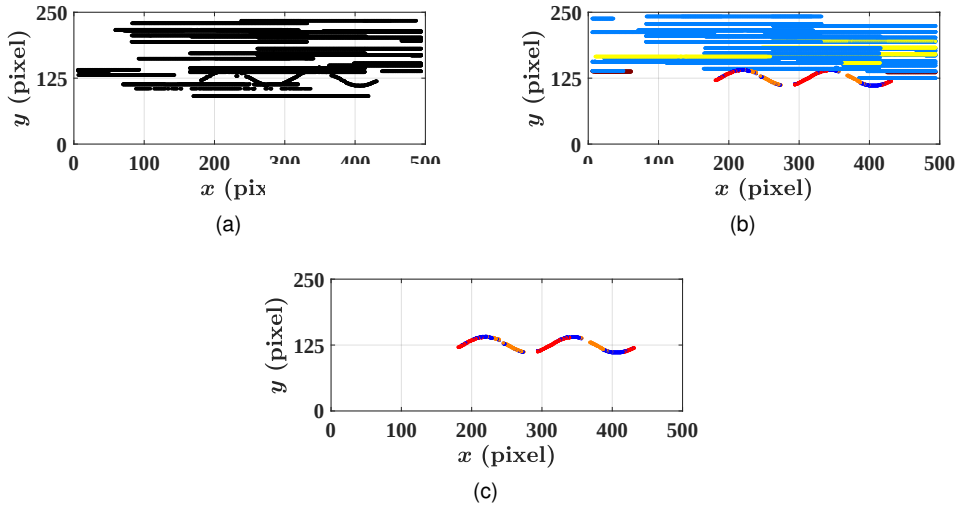


Figure 4.15: Motion traces detected by the ESTMD, DSTMD and STMD+. For fair comparison, the three models have fixed detection rates ($D_R = 0.85$). (a) ESTMD. (b) DSTMD. (c) STMD+. Since the ESTMD cannot detect motion direction, its outputs are all shown in black color.

Fig. 4.15 demonstrates the motion traces detected by the ESTMD, DSTMD and STMD+, where the detection rates (D_R) of the three models are all set to 0.85 for fair comparison. As can be seen, the detection results of the ESTMD and DSTMD are seriously contaminated by a number of fake features, whereas that of the STMD+ is noiseless.

To reveal the role of the contrast pathway, we analyze the directional contrast on two motion traces chosen from Fig. 4.14(a), where one is the small target motion trace, and the other is a randomly selected fake feature trace. Fig. 4.16 presents the directional contrast on these two motion traces. Note that each motion trace has four directional contrast along four directions $\phi \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. As shown in Fig. 4.16(a),

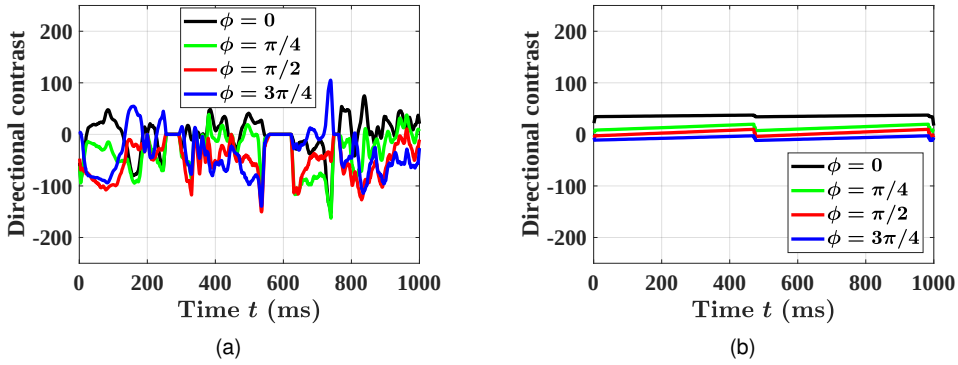


Figure 4.16: (a) Directional contrast on the motion trace caused by the small target. (b) Directional contrast on the motion trace caused by the fake feature. In each subplot, the directional contrast along four directions $\phi \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ is presented.

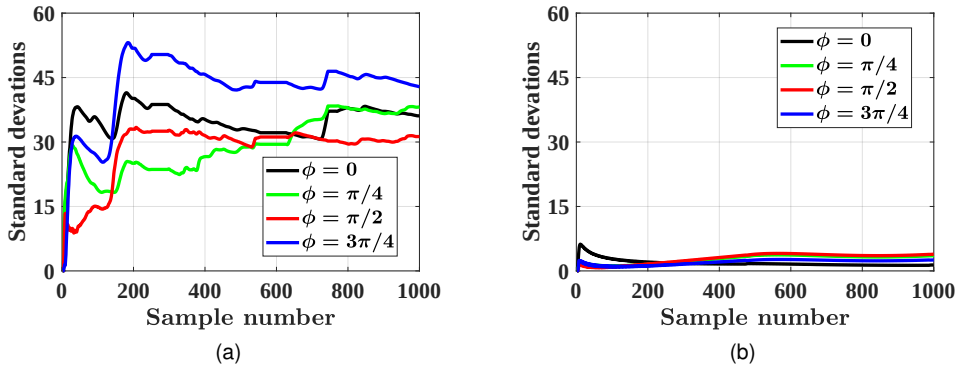


Figure 4.17: Standard deviations under different sample numbers. (a) Standard deviations of the small target. (b) Standard deviations of the fake feature.

the directional contrast on the motion trace caused by the small target displays significant changes over time. In contrast, the directional contrast of the fake feature trace remains almost unchanged with respect to time (see Fig. 4.16(b)). The calculated standard deviations of the directional contrast on these two motion traces are listed in Table 4.3, where the sample number m is equal to 1000. Obviously, the standard deviations of the small target are much larger than those of the fake feature, suggesting that the small target can be discriminated from fake features by comparing their standard deviations.

We further study the relationship of the standard deviations with regard to the sample number m (see Fig. 4.17). As it is shown, the standard deviations of the small target exhibit a sharp rise when the sample number increases from 0 to 200. With the continuous growth of the sample number, the standard deviations tend to be stable. Similarly, the standard deviations of the fake feature become stable as the increase of

Table 4.3: Standard deviations of the direction contrast.

	Standard deviation			
	$\phi = 0$	$\phi = \frac{\pi}{4}$	$\phi = \frac{\pi}{2}$	$\phi = \frac{3\pi}{4}$
Small target	**36.10**	**38.17**	**31.29**	**42.86**
Fake feature	1.38	3.56	3.88	2.54

the sample number. Above results indicate that a certain number of samples is needed to obtain stable standard deviations. Note that the smaller the sample number means the shorter the motion trace, which always results in minor change of the directional contrast. That is, the standard deviation of the small target will be too small to be discriminated from those of fake features when a small sample number is chosen like 50. This is the case in the clear and homogenous backgrounds such as sky and ocean where directional contrast shows little change in a large area. On the other hand, a large sample number needs correctly recording motion traces over a long distance which is extremely difficult to achieve in cluttered backgrounds. Considering these two aspects, the sample number is set as 200 as a tradeoff in the experiments to obtain stable standard deviations.

4.2.4 Comparison on Synthetic and Real Datasets

In this section, six groups of synthetic image sequences are first utilized to evaluate the performance of the proposed model in terms of different target velocities, target sizes, target luminance, background velocities, background motion directions and background images. The details of the synthetic image sequences are listed in Table 4.4. Then the proposed model is further tested on the real dataset (STNS dataset [37]). The performance comparison between the proposed STMD+ and two baseline models (namely, ESTMD and DSTMD), is also conducted.

Fig. 4.18(a) shows the receiver operating characteristics (ROC) curves of the three models for the initial synthetic image sequence. It can be seen that the STMD+ has better performance than the baseline models. More precisely, the STMD+ has higher detection rates (D_R) compared to the baseline models while the false alarm rates (F_A) are low. Fig. 4.18(b)-(d) display the detection rates of the three models for the Group

Table 4.4: Details of the initial image sequence and six groups of image sequences. Comparing to the initial image sequence, Group 1 to 6 are composed of image sequences with different parameters.

Parameter	Initial sequence	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Target velocity (pixel/s)	250	$\mathbf{0 \sim 500}$	250	250	250	250	250
Target size (pixel × pixel)	5×5	5×5	$\mathbf{1 \times 1 \sim 12 \times 12}$	5×5	5×5	5×5	5×5
Target luminance	0	0	0	$\mathbf{0 \sim 75}$	0	0	0
Background velocity (pixel/s)	250	250	250	250	$\mathbf{0 \sim 500}$	$\mathbf{0 \sim 500}$	250
Background motion direction	rightward	rightward	rightward	rightward	rightward	**leftward**	rightward
Background Image	Fig.4.13(a)	Fig.4.13(a)	Fig.4.13(a)	Fig.4.13(a)	Fig.4.13(a)	Fig.4.13(a)	**Fig.4.19(a) ∼ (c)**

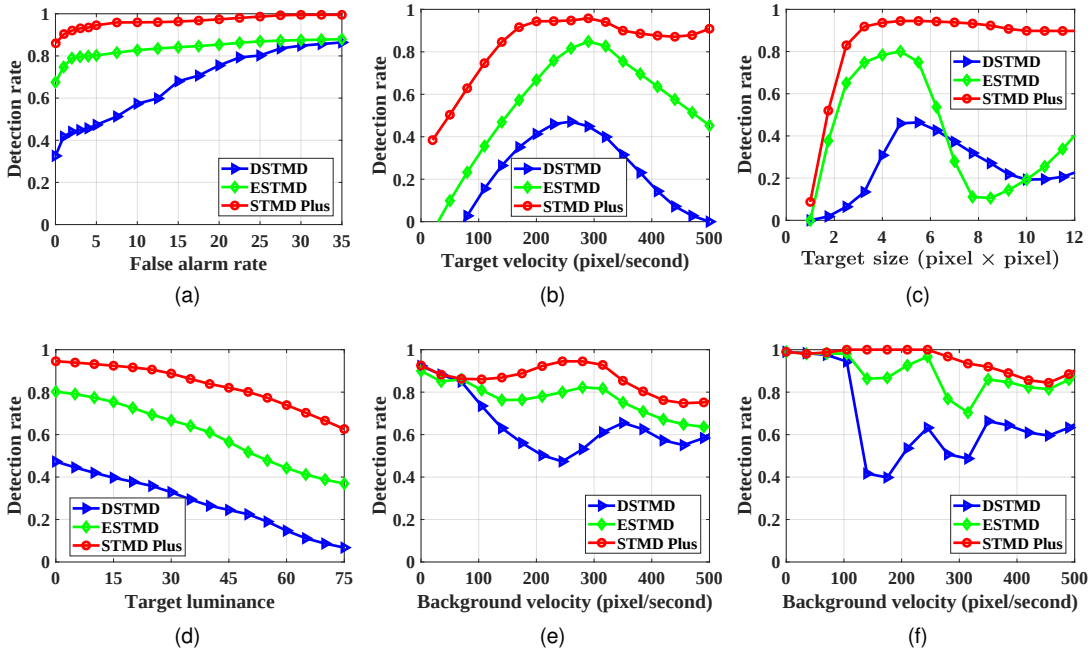


Figure 4.18: (a) Receiver operating characteristic (ROC) curves of the three models for the initial image sequence. (b)-(f) Detection rates of the three models for the Group 1-5. For fair comparison, the three models have fixed false alarm rate ($F_A = 5$). (b) Group 1, different target velocities. (c) Group 2, different target sizes. (d) Group 3, different target luminance. (e) Group 4, different background velocities (in rightward motion). (f) Group 5, different background velocities (in leftward motion).

1 to 5, where the false alarm rates of the three models are all equal to 5 for fair comparison. From Fig. 4.18(b) and (c), we can see that the STMD+ significantly outperforms the baseline models. The STMD+ has higher detection rates than the baseline models for different target velocities and sizes. The detection rate of the STMD+ remains stable when the target velocity (or size) ranges from 200 to 500 pixel/s (or from 4×4 to 12×12 pixel $\times$ pixel). In contrast, the detection rates of the two baseline models significantly decrease after reach the maximum points. As it shown in Fig. 4.18(d), the STMD+ consistently performs best under different target luminance. It is worthy to note that the detection rates of the three models all decrease with the increase of target luminance. In Fig. 4.18(e) and (f), we can see that the STMD+ has the better performance than the baseline models under different background velocities and directions.

Fig. 4.19 presents the ROC curves of the three models for the Group 6. As can be seen, the STMD+ outperforms the baseline models in different backgrounds. Note that the three models all perform well in Fig. 4.19(a). Their detection rates are all close

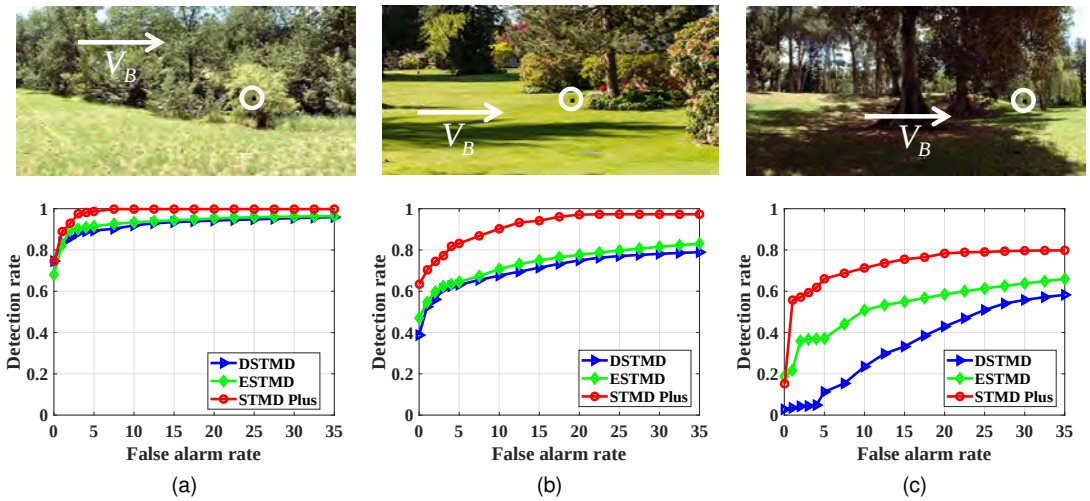


Figure 4.19: Background images and ROC curves of the three models for the Group 6, different backgrounds.

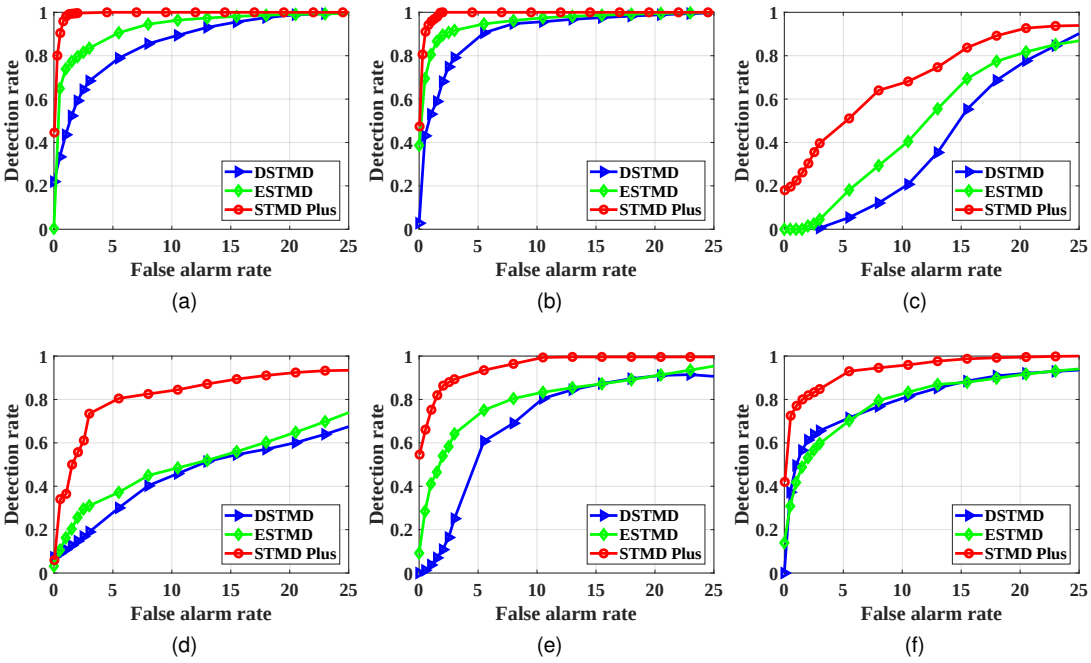


Figure 4.20: ROC curves of the three models for the six real image sequences. (a) Real image sequence 1 (STNS-4). (b) Real image sequence 2 (STNS-15). (c) Real image sequence 3 (STNS-16). (d) Real image sequence 4 (STNS-18). (e) Real image sequence 5 (STNS-22). (f) Real image sequence 6 (STNS-25).

to 1 when the false alarm rates are low, and show small differences. This is because the background is much more homogeneous and contains less fake features. However, in more cluttered backgrounds such as Fig. 4.19(b) and (c), the STMD+ has a much better performance than the other two models.

We further tested the developed model on the publicly available STNS dataset [37]. Fig. 4.20 illustrates the ROC curves of the three models for the six real image se-

quences, where the numbers of these six image sequences in the STNS dataset are 4, 15, 16, 18, 22 and 25, respectively. As it is shown in the six subplots, the detection rates of the STMD+ are higher than those of two baseline models when the false alarm rates are given. That is, the STMD+ obtains the best performance for all six real sequences, which means that the STMD+ can work more stably for different cluttered backgrounds and target types.

4.3 Chapter Summary

In this chapter, we have proposed a visual system model (STMD+) for small target motion detection in cluttered backgrounds. The visual system contains two parallel information pathways and is capable of discriminating small targets from fake features. The first pathway called motion pathway, is intended to locate all small moving objects by calculating luminance changes over time at each pixel. The second pathway called contrast pathway, is designed to capture the directional contrast by computing luminance changes of each pixel along different directions. The mushroom body is introduced to fuse the two types of information from the two pathways. Finally, small targets are distinguished from fake features by comparing the standard deviations of the directional contrast on their motion traces. Comprehensive evaluation on the synthetic and real datasets, and comparisons with the existing STMD models demonstrate the effectiveness of the proposed visual system in filtering out fake features and improving detection rates.

The motion trace recording whose costs scales with $O(p^2)$ in the worst case is so computationally expensive that the STMD Plus may no be suitable for real-time applications. In addition, the performance of the STMD Plus relies on the accuracy of the recorded traces. However, accurately tracking of multiple objects against cluttered backgrounds remains a challenging task due to background clutters, poor contrast, and partial or full occlusions. To avoid high computational complexity and poor performance of motion trace recording, we address the false-positive-elimination problem from the point of feedback mechanisms in the next chapter.

Chapter 5

Feedback STMD – a STMD with Feedback Loop

To discriminate small targets from fake features against cluttered moving backgrounds, Chapter 4 develops the STMD Plus by comparing the changes of directional contrast on the motion traces. It outperforms the other existing STMD-based models in terms of eliminating false positive background features. However, the motion trace recording is computationally expensive, making it inappropriate for real-time applications. Furthermore, the performance of the STMD Plus is limited by the accuracy of the recorded motion traces which is often low in complex environments due to poor contrast, frequent occlusions and complex interactions between objects. To avoid the above limitations, other neural mechanisms, such as feedback, should be integrated with the existing STMD-based models for eliminating false positives instead of the STMD Plus.

Feedback loops, a common neural connection in animals' visual systems, are known to serve a modulatory role in visual processing [244]–[246]. Specifically, they are able to potentiate the abilities of visual neurons to discriminate moving objects from cluttered backgrounds [247]–[250]. Information flowing from higher-level visual areas to lower ones via the feedback loops can effectively enhance responses to the moving objects while suppress those to the background. Although the feedback connections of the STMD pathways are still under investigation, it is clear that incorporating feedback information is beneficial to the improve the performance of the STMD neurons

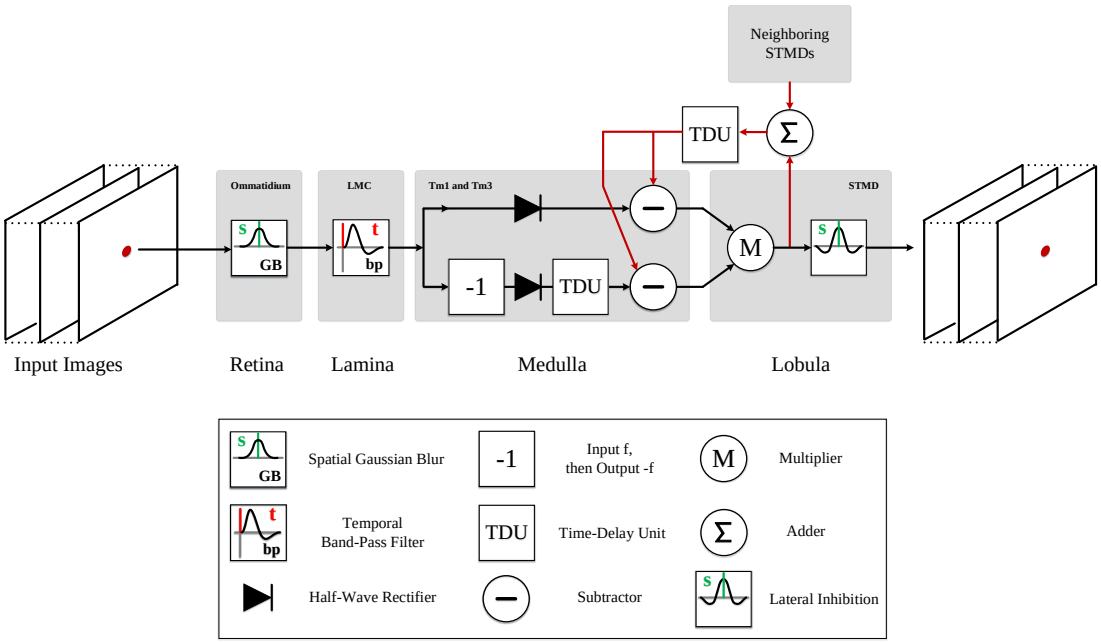


Figure 5.1: Schematic illustration of the proposed feedback STMD-based neural model. For clear illustration, only one STMD is presented here. However, the STMD neurons are arranged in matrix form in the proposed model.

for small target motion detection.

This chapter considers small target discrimination problem from the perspective of feedback mechanisms. A feedback STMD-based neural model named as Feedback STMD, is proposed by applying its model output to the previous medulla layer to construct a feedback loop. The feedback signal is first temporally delayed, then subtracted from the medulla neural outputs to suppress small-target-like background features.

The remainder of this chapter is organized as follows. In Section 5.1, we introduce the developed feedback STMD-based neural model in details. Section 5.2 provides extensive performance evaluation as well as comparisons against the existing models. Finally, we conclude this chapter in Section 5.3.

5.1 Formulation of the Model

The proposed feedback STMD-based neural model is composed of four sequentially arranged neural layers including the retina, lamina, medulla, and lobula, as illustrated in Fig. 5.1. Each layer contains various neurons cooperating together for small target detection. Specifically, luminance signals are received and smoothed by the ommatidia, then propagated to the LMCs to extract luminance changes with respect

to time. The medulla neurons separate the luminance-change signals into increase and decrease components which are further temporally delayed to ensure alignment in temporal domain. These temporally aligned signals are recombined by the STMDs to produce a large response to small moving targets. To construct a feedback loop, the output of the STMD is first delayed, then fed into the previous medulla layer. The formulation of the ommatidia, LMCs, medulla neurons, and STMDs has been elaborated in the previous chapter, so we only describe the proposed feedback loop in the following.

5.1.1 Feedback STMD

As shown in Fig. 5.1, a feedback signal is applied to the medulla neurons to suppress background features by subtracting itself from the medulla neural outputs. After the subtraction, these medulla neural outputs are multiplied together to define the output of the STMDs, that is,

$$\begin{aligned} D_f(x,y,t) = &\Big\{S^{\text{Tm3}}(x,y,t) - \alpha F(x,y,t)\Big\} \\ &\times \Big\{S^{\text{Tm1}}_{(n_3,\tau_3)}(x,y,t) - \alpha F(x,y,t)\Big\} \end{aligned} \tag{5.1}$$

where $D_f(x,y,t)$ stands for the output of the feedback STMDs; $S^{\text{Tm3}}(x,y,t)$ and $S^{\text{Tm1}}_{(n_3,\tau_3)}(x,y,t)$ are medulla neural signals defined in (3.6) and (3.9); $\alpha > 0$ is a constant; $F(x,y,t)$ denotes the feedback signal defined as temporally delayed summation of the STMD neural outputs $D_f(x,y,t)$ and the weighted outputs of surrounding STMDs $E(x,y,t)$, that is

$$F(x,y,t) = \int \Big\{D_f(x,y,s) + E(x,y,s)\Big\}\Gamma_{n_4,\tau_4}(t-s)ds \tag{5.2}$$

where $\Gamma_{n_4,\tau_4}(t)$ denotes a Gamma kernel whose order and time constant are n_4 and τ_4, respectively. The weight function of the surrounding STMDs is given by

$$W_e(x,y) = \frac{1}{2\pi\eta^2}\exp(-\frac{x^2+y^2}{2\eta^2}) \tag{5.3}$$

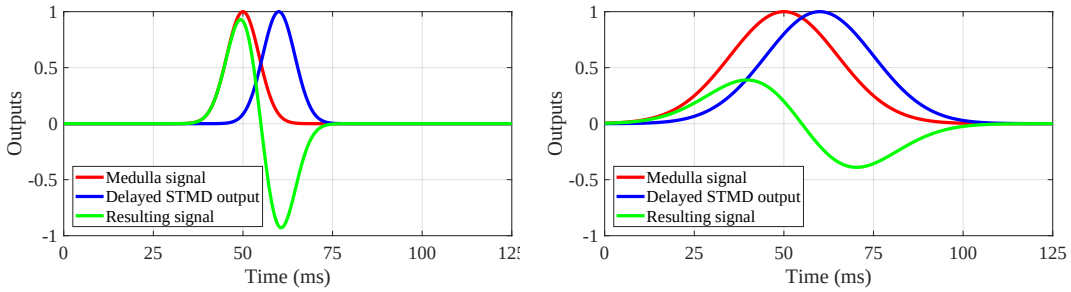


Figure 5.2: Neural outputs at a pixel with time for (a) fast-moving small objects, and (b) moving objects either with low velocities or large sizes. The delay times of the STMD outputs in the two subplots are the same.

where η is constant. Then, the weighted outputs of surrounding STMDs $E(x, y, t)$ is defined as

$$E(x, y, t) = \iint \left\{ S^{\text{Tm3}}(u, v, t) \cdot S^{\text{Tm1}}_{(n_3, \tau_3)}(u, v, t) \right\} W_e(x - u, y - v) du dv. \tag{5.4}$$

Referring to (3.19), the $D_f(x, y, t)$ further convolves with the inhibition kernel $W_s(x, y)$ to suppress large moving objects,

$$E_f(x, y, t) = \iint D_f(u, v, t) W_s(x - u, y - v) du dv \tag{5.5}$$

where $E_f(x, y, t)$ refers to the inhibited output and $W_s(x, y)$ is given in (3.13).

The motivation of the time-delay feedback is based on the following observations: the duration of an object to pass through a pixel is determined by its size and velocity; for objects either with low velocities or with large sizes, their durations are always much longer than those of fast-moving small objects. In the natural world, mates or prey often move faster than the background [122], [251], [252], which means those slow-moving objects are more likely to be background features. As illustrated in Fig. 5.2, the medulla neural outputs with long response duration are largely suppressed after applying the time-delay feedback signal, while those outputs with short response duration are almost preserved. That is, the proposed feedback loop can inhibit slow-moving background features by appropriate time delay.

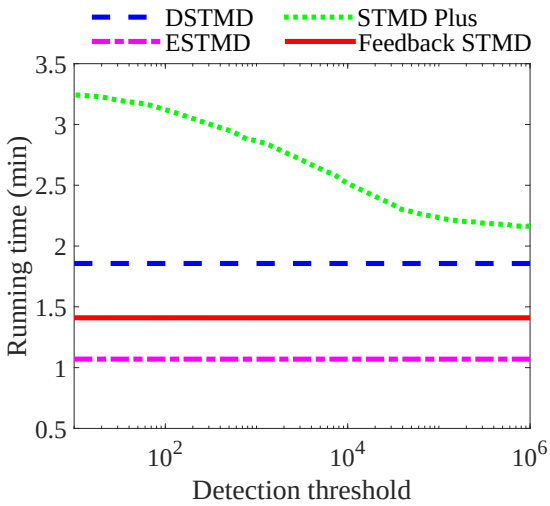


Figure 5.3: Running times of the four models under different detection thresholds for processing 1000 frames.

5.1.2 Computational Complexity

The computational time of the proposed Feedback STMD mainly consists of five parts: the retina, lamina, medulla, lobula and feedback loop, as shown in Fig. 5.1. The complexity of the retina is determined by a 2-D spatial convolution of the input image with a Gaussian kernel, which can be implemented in $O(k^2mn)$ time for an $m \times n$ input image and a $k \times k$ kernel. In the lamina, the LMC output can be regarded as the difference of two Gamma convolutions. Since the temporal Gamma convolution needs $O(lmn)$ cost where l is the length of the Gamma kernel, the computational complexity of the LMC scales with $O(2lmn)$. Two medulla neural outputs are multiplied together to define the STMD output, so the total cost of this step is about $O(lmn + 2mn)$ where the complexity for calculating the two medulla neural outputs is $O(lmn + mn)$. Finally, the lateral inhibition mechanism which is implemented by a 2-D convolution, needs $O(k^2mn)$ cost. In the feedback loop, the cost for calculating the temporally delayed feedback signal grows like $O(lmn)$.

Based on the above analysis, the entire computational complexity of the proposed Feedback STMD is around $O(N(2k^2 + 4l + 2)mn)$ where N stands for the number of input images. To quantitatively compare the complexity of the existing STMD-based models, we record their running times for processing 1000 frames under different detection thresholds. As can be seen from Fig. 5.3, the running times of the DSTMD, ESTMD, and Feedback STMD remain almost unchanged under different detection

Table 5.1: Running times of the four models for processing 1000 frames where the detection threshold is set as 10.

	ESTMD	Feedback STMD	DSTMD	STMD Plus
Time (min)	1.07	1.41	1.86	3.12

Table 5.2: Parameters of the proposed Feedback STMD

Eq.	Parameters
(3.1)	$\sigma_1 = 1$
(3.3)	$n_1 = 4, \tau_1 = 8, n_2 = 16, \tau_2 = 32$
(5.1)	$n_3 = 9, \tau_3 = 45, \alpha = 1$
(5.2)	$n_4 = 10, \tau_4 = 25$
(5.5)	$A = 1, B = 3, \sigma_2 = 1.5, \sigma_3 = 3, e = 1, \rho = 0$

thresholds. However, the running time of the STMD Plus significantly increases as the decrease of the detection threshold. This is because the complexity of the STMD Plus for recording motion traces scales with $O(p^2)$ where p is the number of detected objects. The lower detection threshold always means the higher the number of detected objects p, which finally results in the non-linear growth of the running time. The running times of the four models when the detection threshold is set as 10 are given in Table 5.1. As it is shown, the running time of the Feedback STMD is much less than that of the STMD Plus, which approximates the half of the STMD Plus's running time.

5.1.3 Parameter Setting

Parameters of the proposed Feedback STMD are listed in Table 5.2, where the parameters of four neural layers are determined by the analysis in [132], [133]. The feedback constant α and time-delay length τ_4 will be analysed to evaluate their effects on the model performance in Section 5.2.3. On the whole, all parameters are chosen to satisfy the basic neural properties which are mainly determined by velocity and size ranges of moving targets.

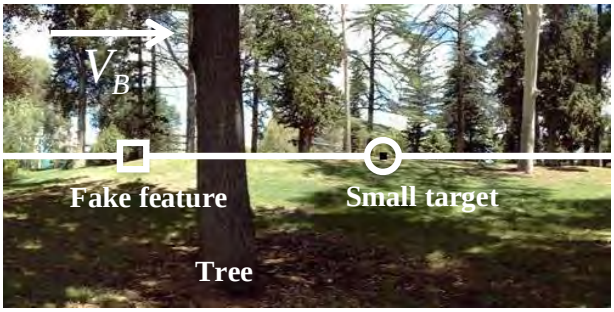


Figure 5.4: Input frame at time $t_0 = 1000$ ms whose resolution is 500 pixels (in horizontal) by 250 pixels (in vertical). The small target (the black block) is moving against the cluttered background where arrow V_T and V_B denote the motion direction of the target and background, respectively. The velocity of the small target and background are set as 250 pixel/s and 150 pixel/s, respectively. The fake feature is embedded in the background and also moving due to background motion.

5.2 Results and Discussions

The Vision Egg dataset [225] is used to verify the proposed feedback neural model. It includes various image sequences, each of which is synthesized by real background images and computer generated small targets. These image sequences all display a small target moving against cluttered backgrounds, but differ in target size, target velocity, target luminance, background velocity, background types and so on. The temporal sampling frequency of the videos is 1000 Hz.

5.2.1 Signal Processing of Various Neurons

To intuitively illustrate the signal processing of the developed model, we observe the output of each neural layer with respect to x by setting y and t as $y_0 = 125$ pixel and $t_0 = 1000$ ms. Fig. 5.4 displays the input frame at time $t_0 = 1000$ ms where the luminance signal $I(x, y_0, t_0)$ on the middle line is shown in Fig. 5.5(a). The resulting ommatidium output and LMC output are presented in Fig. 5.5(b) and (c), respectively. As can be seen, the ommatidium output is slightly smoothed compared to the input signal; the LMC output is a temporal band-pass filtered version of the ommatidium output where positive LMC outputs mean luminance increase of pixels while negative LMC outputs suggest luminance decrease over time.

To reveal the role of the feedback loop, we compare the inputs and outputs of the STMD with those of the Feedback STMD. Fig. 5.6(a) demonstrates the two in-

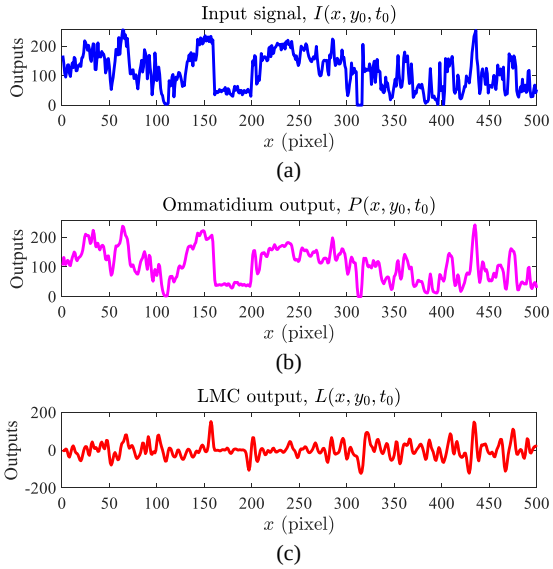


Figure 5.5: In each subplot, the horizontal axis denotes x coordinate while the vertical axis represents neural outputs. (a) Input luminance signal. (b) Ommatidium output. (c) LMC output.

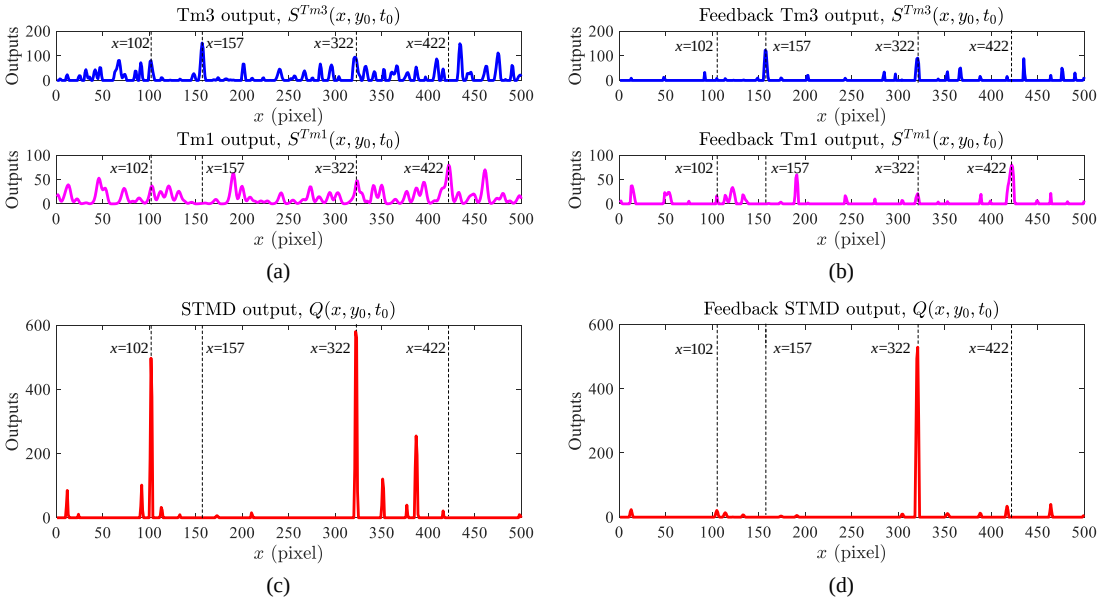


Figure 5.6: In each subplot, the horizontal axis denotes x coordinate while the vertical axis represents neural outputs. (a) Two inputs of the STMD. (b) Two inputs of the Feedback STMD. (c) Output of the STMD. (d) Output of the Feedback STMD.

puts of the STMD where $S^{\text{Tm3}}(x, y_0, t_0)$ is the positive part of the LMC output while $S^{\text{Tm1}}_{(n_3,\tau_3)}(x, y_0, t_0)$ represents the negative part of the LMC output with time delay. These two medulla signals are multiplied together then laterally inhibited to produce strong responses to small moving targets. As can be seen from Fig. 5.6(c), a large STMD response appears at the location of the small target ($x = 322$), however, the other positions such as the location of the fake feature $x = 102$, also present strong STMD

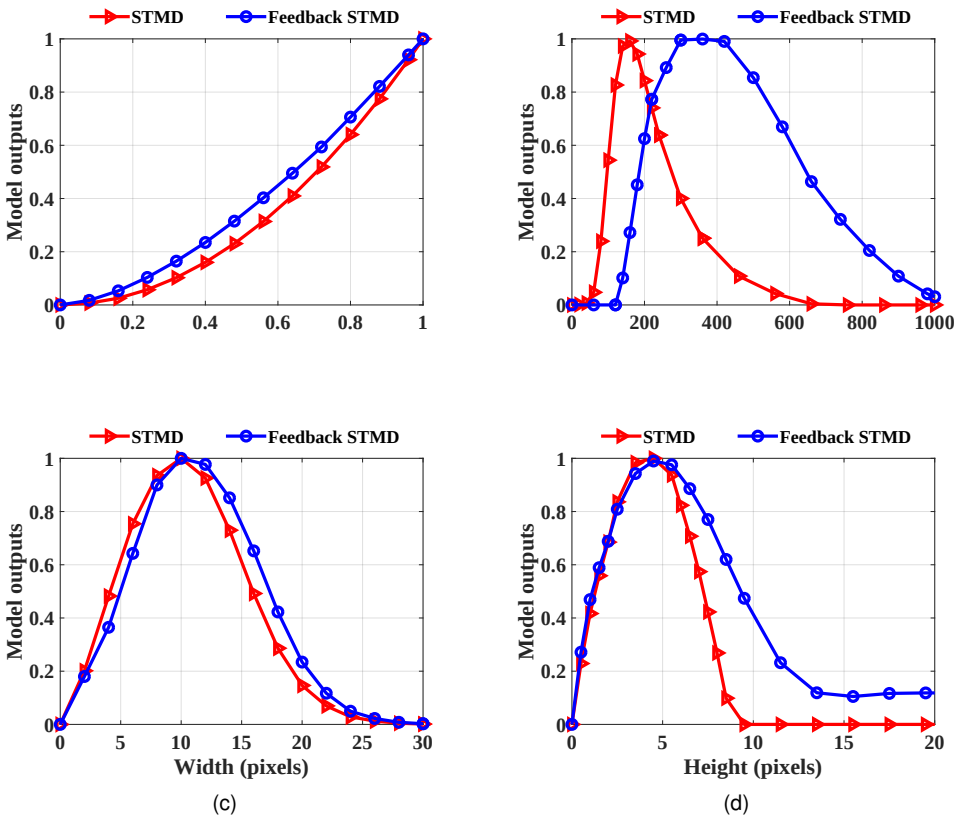


Figure 5.7: Outputs of the STMD and Feedback STMD to moving objects with different Weber contrast, velocities, widths and heights. (a) Different Weber contrast. (b) Different velocities. (c) Different widths. (d) Different heights.

responses. To suppress the responses to the background features, a feedback signal is applied to the medulla neurons. Fig. 5.6(b) and (d) illustrates the two medulla neural outputs after feedback and the output of the feedback STMD, respectively. As it is shown, the responses to the background features are largely suppressed after the feedback, whereas the response to the small target still maintains a much higher value. In this case, the small target can be easily distinguished by comparing the model output with a preset threshold.

5.2.2 Tuning Properties

The STMD model exhibits four tuning properties including Weber contrast sensitivity, velocity selectivity, width selectivity and height selectivity. However, whether the STMD with a feedback loop can preserve these four properties remains unclear. To answer this question, we compare the outputs of the feedback STMD to objects with different velocities, widths, heights and Weber contrast. The initial Weber contrast,

velocity, width and height of the object are set as 1, 250 pixel/s, 5 pixels and 5 pixels.

Fig. 5.7(a) shows the outputs of the STMD and feedback STMD with respect to Weber contrast. As can be seen, both models exhibit Weber contrast sensitivity. Specifically, their outputs increase as the increase of Weber contrast, until reaches maximum at Weber contrast $= 1$. Fig. 5.7(b) displays the outputs of the STMD and feedback STMD with regard to object velocity. As it is shown, both models are selective to the object's velocity. Their outputs all peak at respective optimized velocities, where the optimized velocity of the feedback STMD (400 pixel/s) is much higher than that of the STMD (150 pixel/s). In addition, the feedback STMD exhibits high responses to the objects whose velocities range from 150 pixel/s to 1000 pixel/s, while the response range of the STMD is between 50 pixel/s and 600 pixel/s. Fig. 5.7(c) presents the outputs of the STMD and feedback STMD to objects with different widths, indicating that both models prefers moving objects whose widths are smaller than 25 pixels. Fig. 5.7(d) illustrates the outputs of the two models with respect to object height. As can be seen, both models' outputs reach their maximum at an optimized height $= 5$ pixels. When object height is larger than 15 pixels, the output of the feedback STMD significantly decreases and remains stable at 0.1, whereas that of the STMD is close to 0. In summary, the above results reveal that the developed feedback STMD preserves the four tuning properties.

5.2.3 Parameter Sensitivity

Feedback constant α and temporal delay length τ_4 are two important parameters to determine the feedback signal, as can be seen from Equation (5.1) and (5.2). To evaluate the impacts of these two parameters, we compare the tuning properties of the developed Feedback STMD under different α and τ_4, while other parameters are kept at their initial values in Table 5.2.

Fig. 5.8 presents the four tuning properties under different feedback constants α. As can be seen from Fig. 5.8(a) and (c), the increase of the feedback constant α cannot induce significant changes of tuning curves, which means that it has little effect on Weber contrast sensitivity and width selectivity. In contrast, the optimal velocity will

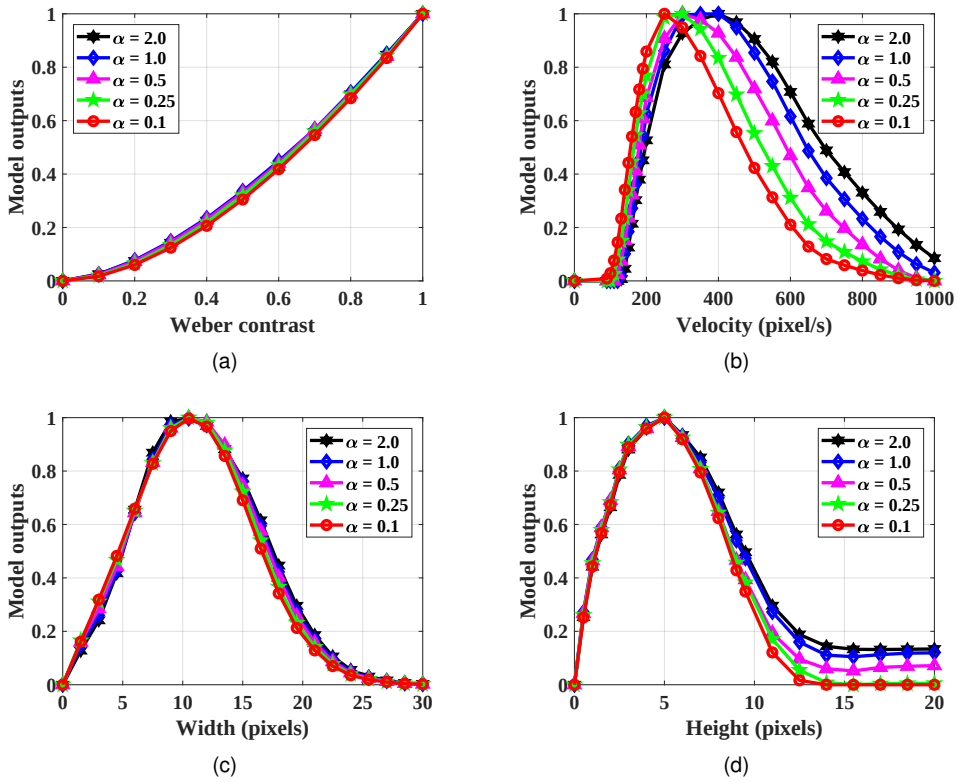


Figure 5.8: Tuning properties of the Feedback STMD under different feedback constants α which are set as 2.0, 1.0, 0.5, 0.25 and 0.1, respectively. (a) Weber contrast sensitivity. (b) Velocity selectivity. (c) Width selectivity. (d) Height selectivity.

increase with the increase of the α as shown in Fig. 5.8(b). Meanwhile, the preferred velocity range will also shift to higher values, indicating that the moving objects with high velocities (> 250 pixel/s) can more easily be detected at the larger α. In Fig. 5.8(d), we can see that the feedback constant α has little impact on the optimal height which remains at 5 pixels for different α. In addition, the responses to objects whose heights are larger than 10 pixels will have slight growth as the increase of α, but the preferred height range (from 0 to 10 pixels) is unchanged. Fig. 5.9 demonstrates the four tuning properties under different time-delay length τ_4. It can be seen that Weber contrast sensitivity, width selectivity and height selectivity are little affected by τ_4; on the contrary, the optimal velocity will decrease as the increase of the time-delay length.

The reason for the above results is - feedback constant α and time-delay length τ_4 determine the strength of the feedback signal. More precisely, the higher α (or the lower τ_4) always means the stronger feedback signal. Since the feedback signal is the delayed version of the STMD model output, it contains motion information about the slowly moving objects which are always background fake features. When the larger

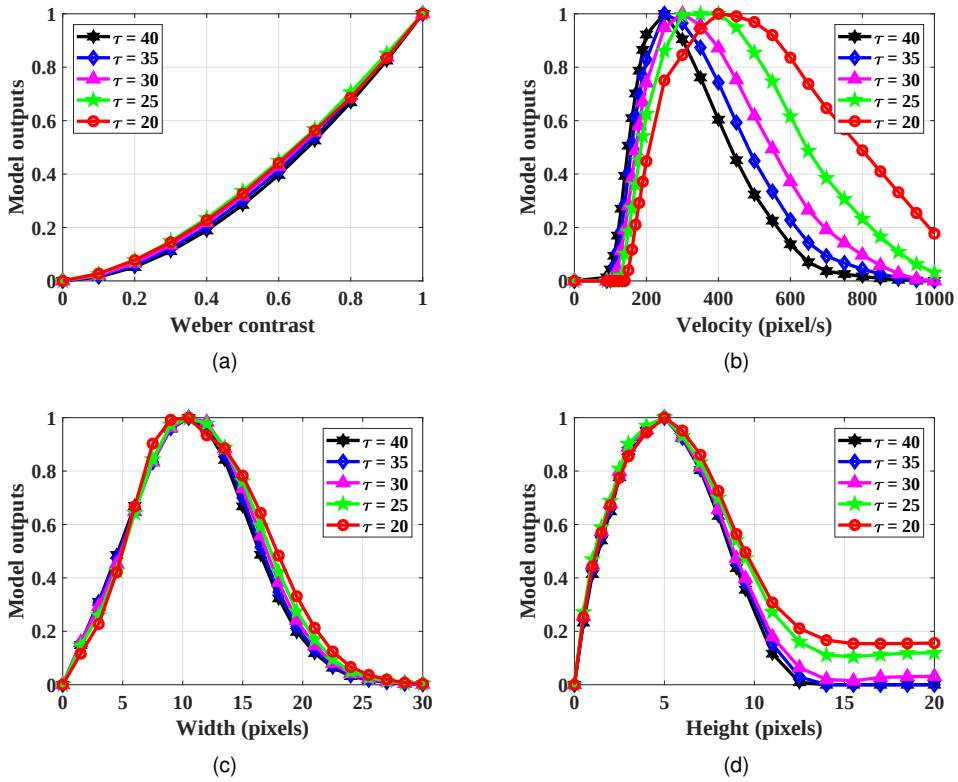
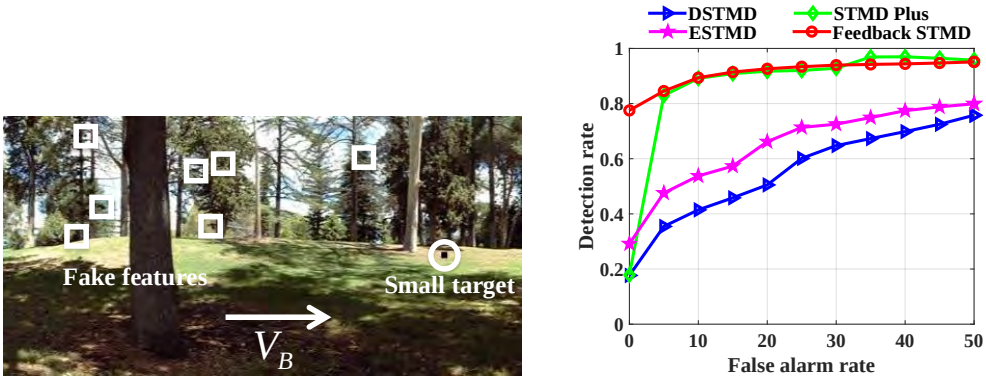


Figure 5.9: Tuning properties of the feedback STMD under different feedback time-delay length τ_4 which are set as 40, 35, 30, 25 and 20, respectively. (a) Weber contrast sensitivity. (b) Velocity selectivity. (c) Width selectivity. (d) Height selectivity.

feedback signal is applied to the previous medulla layer, the responses to those slowly moving fake features will be much weaker. That is, the Feedback STMD model prefers moving objects with higher velocities.

5.2.4 Comparison on the Vision Egg Dataset

In this section, a synthetic image sequence is first utilized to evaluate the performance of the proposed Feedback STMD model. The performance comparison between the Feedback STMD and three baseline models (namely, the ESTMD, DSTMD and STMD Plus), is also conducted. For a fair comparison, the parameters of the three baseline models are all tuned to ensure they have the same preferred velocity and size ranges as the STMD in Fig. 5.7. As can be seen from Fig. 5.10(a), the input image sequence displays the motion of a small target against the cluttered background where its parameters are listed in Table 5.3. The receiver operating characteristics (ROC) curves of the four models for the initial synthetic image sequence are shown in Fig. 5.10(b). It can be seen that the Feedback STMD achieves the same performance as the STMD



(a) (b)

Figure 5.10: (a) Representative frame of the initial image sequence. The small target (the small black block) highlighted by the circle, is moving against the cluttered background. The background which contains a number of fake features, is also moving from left to right where arrow V_B denotes the background motion direction. (b) Representative operating characteristic (ROC) curves of the three models for the initial image sequence.

Table 5.3: Parameters of the initial image sequence.

Parameter	Initial sequence
Target velocity (pixel/s)	250
Target size (pixel × pixel)	5×5
Target luminance	0
Background velocity (pixel/s)	150
Background motion direction	rightward

Plus while consuming much less computational resources. In addition, they all clearly outperform the DSTMD and ESTMD. More precisely, the Feedback STMD has higher detection rates (D_R) compared to the two models while the false alarm rates (F_A) are low.

We further change the parameters of the input image sequences to demonstrate the performance of the Feedback STMD in terms of different target velocities, target sizes, target luminance, background velocities, background motion directions. Fig. 5.11(a)-(e) presents the detection rates of the four models for image sequences with different parameters, where the false alarm rates of the four models are all equal to 10 for fair comparison. As shown in Fig. 5.11(a), the Feedback STMD consistently performs best under different target luminance. It is also worthy to mention that the detection rates of the models all decrease as the increase of target luminance. From Fig.

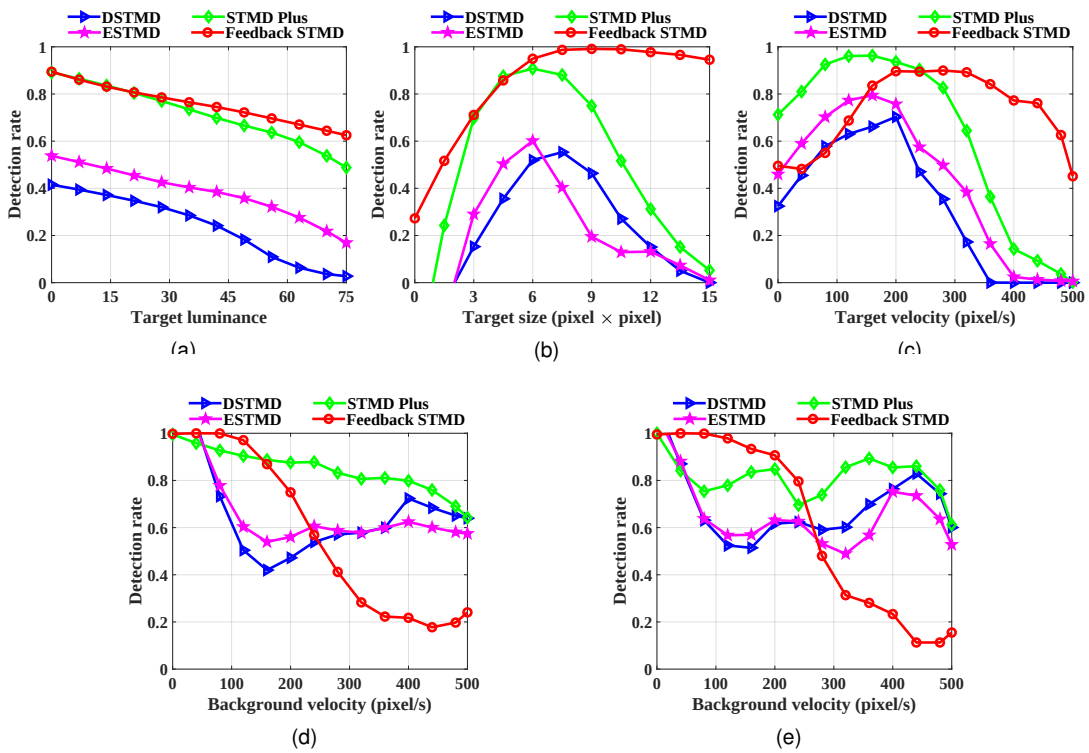


Figure 5.11: Detection rates of the three models for image sequences with different parameters. For fair comparison, the three models have fixed false alarm rate ($F_A = 10$). (a) Different target luminance. (b) Different target sizes. (c) Different target velocities. (d) Different background velocities (in rightward motion). (e) Different background velocities (in leftward motion).

5.11(b), we can see that the Feedback STMD significantly outperforms the baseline models. The Feedback STMD has higher detection rates than the baseline models for different target sizes. Moreover, the detection rates of the three baseline models are all close to 0 for the target whose size is larger than 20×20 pixels, while that of the Feedback STMD is still higher than 0.9. As can be seen from Fig. 5.11(c), the Feedback STMD has much higher detection rate than the three baseline methods when the target velocity is larger than 250 pixel/s. Similarly, in Fig. 5.11(d) and (e), the Feedback STMD significantly improves the detection rate when the background velocity is lower than the target velocity (250 pixel/s). The above results indicate that the developed Feedback STMD prefers fast-moving small targets and can achieve better performance when the background velocity is lower than the target velocity.

Fig. 5.12 presents the ROC curves of the four models for different backgrounds. As can be seen, the Feedback STMD outperforms the baseline models in different backgrounds. Note that the performances of the models in Fig. 5.12(a) are all better

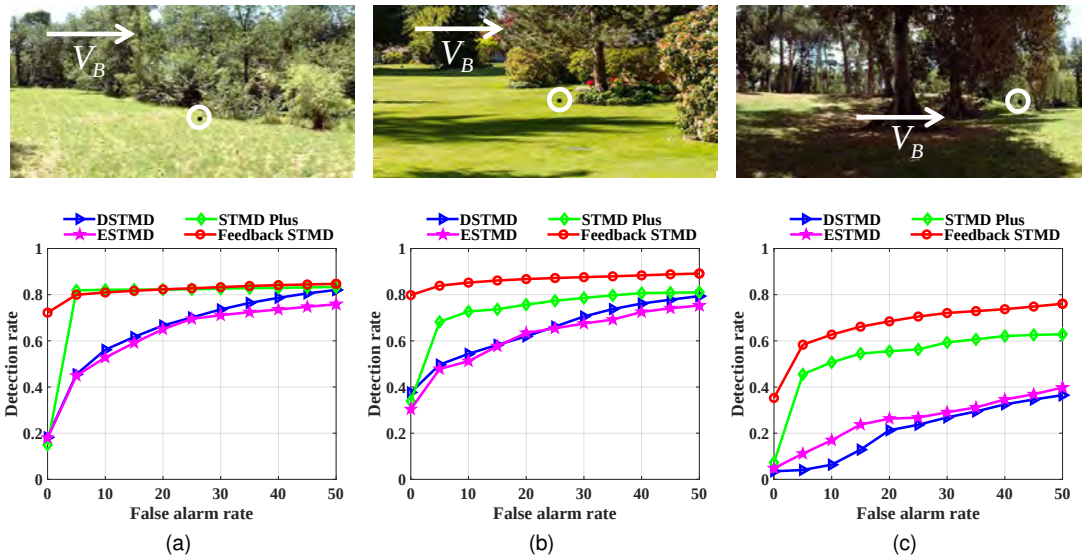


Figure 5.12: Background images and ROC curves of the three models for image sequences with different backgrounds.

than those in Fig. 5.12(c). This is because the background shown in Fig. 5.12(a) is much more homogeneous and contains less fake features. In this case, the detection rates of the four models are quite high when the false alarm rates are low.

5.3 Chapter Summary

In this chapter, we have developed a feedback STMD-based neural model for small target motion detection. The Feedback STMD contains four neural layers and a feedback loop, which is able to discriminate small targets from fake features. The four neural layers are designed to detect small moving objects in the form of producing large responses by calculating temporal luminance changes at each pixel. The feedback loop is intended to suppress background features by applying the temporally delayed neural output to the previous neural layer. Experimental results show that the proposed feedback neural model achieves better performance than the existing STMD-based models in discriminating small targets from complex backgrounds.

Chapter 6

Research Contributions and Future Work

In this chapter, we summarize the main points made in Chapters $3-5$ and point out the future directions that can be carried out to extend the research of this thesis.

6.1 Research Contributions

This dissertation explores three novel STMD-based neural models to detect small target motion against cluttered moving backgrounds. It simulates the basic characteristics of the STMD neurons, provides rigorous mathematical description and systematically tests the developed neural models. The main contributions of the dissertation are summarized below.

- A directionally selective model (DSTMD) is designed to simulate the specific STMD neuron with directional selectivity. It employs a new correlation mechanism which correlates signals from two different pixels to introduce directional selectivity. A lateral inhibition mechanism implemented on the correlation output is proposed for size selectivity. To estimate motion directions of small targets, a population vector algorithm is developed. The experimental results show that the DSTMD can detect not only small moving targets, but also their motion directions against complex backgrounds.

- A visual system model (STMD Plus) is proposed to discriminate small targets from background fake features. To locate small moving objects, a motion pathway which calculates luminance changes over time at each pixel is designed. Additionally, a contrast pathway is developed to capture the directional contrast by computing luminance changes of each pixel along different directions. Finally, the mushroom body is designed to integrate motion information and directional contrast from the two pathways. Small targets are distinguished from fake features by comparing the standard deviations of the directional contrast on their motion traces. Experimental results demonstrate the STMD+ outperforms the existing STMD-based models in eliminating fake features.

- A feedback STMD-based neural model (Feedback STMD) is also developed to distinguish small targets from background fake features. It contains four neural layers and a feedback loop so as to avoid expensive computational cost for recording motion traces. The four neural layers are proposed to detect small moving objects by calculating luminance changes with respect to time. To suppress background features, a feedback loop is designed to propagate the time-delay model outputs to the previous neural layer. Experimental results show that the Feedback STMD achieves better performance than the existing STMD models in inhibiting background fake features while consuming much lower computational resources.

The STMD-based neural models presented in this dissertation demonstrated reliable abilities to detect small moving targets against complex backgrounds. The parameters of the developed models are mainly determined by the velocity and size of the preferred targets. If the preferred target velocity v_0, width w_0, and height h_0 are given, the delay length τ_4, τ_5, and τ_6 can be decided by $\tau_4 = \frac{\alpha}{v_0}$, $\tau_5 = \frac{w_0}{v_0}$, and $\tau_6 = \tau_4 + \tau_5$, while the σ_2 and σ_3 of the lateral inhibition kernel can be given by $\sigma_2 = \frac{h_0}{2\sqrt{2}}$ and $\sigma_3 = 2\sigma_2$ (see (3.13), (3.18), and the discussion in Section 3.2.3). The above parameters are important for the developed models to ensure velocity and size selectivities. In addition, the tuning curves of the models will peak at v_0, w_0 and h_0 (see Fig. 3.12). The other parameters are slightly adjusted to satisfy the functionality based on the

aforementioned parameters.

Visual pursuit of small objects is an important task to be solved by many animals, which requires the target to be detected, then to be fixated in the centre of the visual field and finally to be followed by appropriate movements of the eyes, the head and/or the entire body [252]. However, the movement towards a target will lead to a displacement of the entire retinal image of the environment in the opposite direction. This wide-field motion may evoke an optomotor following response that drives the animal to turn away from the moving target. Although the pursuit of small objects and following responses to wide-field motion seem to be in conflict with each other, the insects' visual systems are able to effectively coordinate these two visually driven behaviors together. Nowadays, visual sensors are becoming more reliable while the computational ability of the processors is more powerful. These make it possible for mobile robots such as unmanned aerial vehicle (UAV) equipped with visual sensors and the presented neural models to detect small moving targets in the distance in the real world. However, the interaction between the small target detection system and the flight control system would be a challenging problem for the embodiment on robots. The insects' visual systems have elegantly addressed this issue, which could provide effective solutions for robots to mediate the two systems.

The proposed models take inspirations from animals' visual systems, but they could also provide possible directions for further investigating the STMD neuron and its neural circuits. For example, the Feedback STMD proposed in Chapter 5 offers a possible explanation for the existence of the feedback loop in the visual pathway of STMD neurons. The Feedback STMD satisfies the basic properties of the STMD neurons, but also shows the preference for the moving objects whose velocities are higher than those of the background. The preference for fast-moving objects is in accord with the biological findings on animals' visual systems [122], [251], [252], indicating that a feedback loop could be a possible factor for generating such preference. That is, the developed Feedback STMD model opens up the possibilities of the existence of the feedback loop in the STMD visual pathways, which could be demonstrated by further biological experiments. By designing artificial visual systems, testing the system'

properties, and comparing them with biological findings, biologists could not only have a further understanding of biological visual systems, but also predict neural properties and circuits.

6.2 Future Work

Animals' visual systems provide a rich source of inspiration for designing artificial visual systems to detect small moving targets against complex dynamic environments. The studies discussed in this thesis represent only a small step in modelling the insects' visual systems for small target motion detection. The following research directions are interesting to further explore in the future.

- Investigating the self-adaptability of the developed neural models. In this dissertation, the parameters of the neural models are all tuned to satisfy the basic properties of the STMD neurons, which are mainly determined by preferred target sizes and velocities. However, the sizes and velocities of target are changing constantly during the visual pursuit. To obtain a better detection performance, the models' parameters should be adjusted to fit these changes. In addition, the detection threshold of the models are fixed in different environments such as sky and bushes whose background clutters always vary significantly. A self-adaptable detection threshold will contribute to the improvement of detection rates. To address these issues, optimization algorithms like genetic algorithm [25], [253], [254] may provide a possible way to obtain the optimal model parameters for various targets and environments. More precisely, a measurement to evaluate whether the parameters are optimal for a specific target and/or background needs to be first constructed, then the optimal parameter searching can be transformed into a optimization problem which can be solved by the genetic algorithms.

- Modelling other neural circuits or mechanisms, and integrating them with the existing STMD-based models. In the insects' visual systems, multiple neural circuits are coordinated to discriminate small moving targets. The neural mod-

els proposed in this dissertation represents only a small part of the insects' visual systems, which could achieve more robust detection performance or more functionalities by integrating with other visual circuits. For example, the embodiment of the developed small target detecting models on robots will require the cooperation between different systems. Specifically, a wide-field motion detection system is needed to estimate background motion for flight control of the UAVs, while the small target motion detection system should be further improved to accommodate the unstability of the moving platforms. In addition, the interactions between these two systems would be a challenging problem, but are important for visual pursuit of small targets in flight. Further effort can be made to investigate the cooperation of visual circuits in the insects' visual systems so as to design an effective interaction scheme for the moving robots.

- Exploring and employing image statistics of natural scenes for small target detection. Natural scenes appear to be extremely complex, and to provide redundant information for the motion detection. However, the insects have shown excellent ability to reduce redundant visual information in natural environments. As revealed in [255], the natural scenes have specific constraints in spatio-temporal space, which means that their image statistics such as color, contrast and entropy could be important cues for insects to optimize visual information. Once the statistics of the natural scenes are identified, they can be used in the early processing step of the STMD-based models to remove most of background features and roughly locate small moving targets.

Bibliography

[1] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for uav tracking," in *Proc. Conf. ECCV*, pp. 445–461, Springer, 2016.

[2] T. J. Wardill, K. Knowles, L. Barlow, G. Tapia, K. Nordström, R. M. Olberg, and P. T. Gonzalez-Bellido, "The killer fly hunger games: target size and speed predict decision to pursuit," *Brain Behav. Evol.*, vol. 86, pp. 28–37, Sep. 2015.

[3] A. Borst, "Fly visual course control: behaviour, algorithms and circuits," *Nat. Rev. Neurosci.*, vol. 15, pp. 590–599, Aug. 2014.

[4] J. R. Sanes and S. L. Zipursky, "Design principles of insect and vertebrate visual systems," *Neuron*, vol. 66, pp. 15–36, Apr. 2010.

[5] P. B. Cummings, *Lecture Presentations for Biology Eighth Edition Neil Campbell*, 2008. [Online]. Available: https://slideplayer.com/slide/6966819/.

[6] A. Borst and M. Helmstaedter, "Common circuit design in fly and mammalian motion vision," *Nat. Neurosci.*, vol. 18, pp. 1067–1076, Jun. 2015.

[7] R. Behnia, D. A. Clark, A. G. Carter, T. R. Clandinin, and C. Desplan, "Processing properties of on and off pathways for drosophila motion detection," *Nature*, vol. 512, pp. 427–430, Aug. 2014.

[8] J. Carbone, A. Yabo, and D. Oliva, "Characterization and modelling of looming-sensitive neurons in the crab neohelice," *J. Comp. Physiol. A*, pp. 1–17, 2018.

[9] R. B. Dewell and F. Gabbiani, "Biophysics of object segmentation in a collision-detecting neuron," *ELife*, vol. 7, Apr. 2018, Art. no. e34238.

[10] S. Yue and F. C. Rind, "Collision detection in complex dynamic scenes using an lgmd-based visual neural network with feature enhancement," *IEEE Trans. Neural Netw.*, vol. 17, pp. 705–716, May 2006.

[11] C. Hu, F. Arvin, C. Xiong, and S. Yue, "Bio-inspired embedded vision system for autonomous micro-robots: the lgmd case," *IEEE Trans. Cogn. Develop. Syst.*, vol. 9, pp. 241–254, Sep. 2016.

[12] H. Eichner, M. Joesch, B. Schnell, D. F. Reiff, and A. Borst, "Internal structure of the fly elementary motion detector," *Neuron*, vol. 70, pp. 1155–1164, Jun. 2011.

[13] P. D. Barnett, K. Nordström, and D. C. O'Carroll, "Retinotopic organization of small-field-target-detecting neurons in the insect visual system," *Curr. Biol.*, vol. 17, pp. 569–578, Apr. 2007.

[14] D. M. Bolzon, K. Nordström, and D. C. O'Carroll, "Local and large-range inhibition in feature detection," *J. Neurosci.*, vol. 29, pp. 14143–14150, Nov. 2009.

[15] K. Nordström and D. C. OCarroll, "Feature detection and the hypercomplex property in insects," *Trends Neurosci.*, vol. 32, pp. 383–391, Jun. 2009.

[16] D. Tolhurst, Y. Tadmor, and T. Chao, "Amplitude spectra of natural images," *Ophthal. Physiol. Opt.*, vol. 12, pp. 229–232, Apr. 1992.

[17] K. Nordström, P. D. Barnett, and D. C. O'Carroll, "Insect detection of small targets moving in visual clutter," *PLoS Biol.*, vol. 4, Feb. 2006, Art. no. e54.

[18] B. R. Geurten, K. Nordström, J. D. Sprayberry, D. M. Bolzon, and D. C. O'Carroll, "Neural mechanisms underlying target detection in a dragonfly centrifugal neuron," *J. Exp. Biol.*, vol. 210, no. 18, pp. 3277–3284, 2007.

[19] M. Egelhaaf, "On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly," *Biol. Cybern.*, vol. 52, pp. 123–140, Jun. 1985.

[20] P. Hennig and M. Egelhaaf, "Neuronal encoding of object and distance information: a model simulation study on naturalistic optic flow processing," *Front. Neural Circuits*, vol. 6, Mar. 2012, Art. no. 14.

[21] P. Hennig, R. Möller, and M. Egelhaaf, "Distributed dendritic processing facilitates object detection: a computational analysis on the visual system of the fly," *PLoS One*, vol. 3, pp. 1–14, Aug. 2008.

[22] M. Zhao, J. Zhao, S. Zhao, and Y. Wang, "A novel method for moving object detection in intelligent video surveillance systems," in *Int. Conf. Comput. Intell. Sec.*, vol. 2, pp. 1797–1800, IEEE, 2006.

[23] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE Trans. Image Process.*, vol. 22, pp. 4996–5009, Sep. 2013.

[24] S. Yue, F. C. Rind, M. S. Keil, J. Cuadri, and R. Stafford, "A bio-inspired visual collision detection mechanism for cars: Optimisation of a model of a locust neuron to a novel environment," *Neurocomputing*, vol. 69, pp. 1591–1598, Feb. 2006.

[25] S. Yue and F. C. Rind, "Redundant neural vision systems-competing for collision recognition roles," *IEEE Trans. Auton. Ment. Dev.*, vol. 5, pp. 173–186, Apr. 2013.

[26] R. Olberg, A. Worthington, and K. Venator, "Prey pursuit and interception in dragonflies," *J. Comp. Physiol. A*, vol. 186, pp. 155–162, Feb. 2000.

[27] M. Mischiati, H.-T. Lin, P. Herold, E. Imler, R. Olberg, and A. Leonardo, "Internal models direct dragonfly interception steering," *Nature*, vol. 517, pp. 333–338, Jan. 2015.

[28] K. Nordström, "Neural specializations for small target detection in insects," *Curr. Opin. Neurobiol.*, vol. 22, pp. 272–278, Apr. 2012.

[29] K. Nordström and D. C. O'Carroll, "Small object detection neurons in female hoverflies," *Proc. Royal Soc. B*, vol. 273, pp. 1211–1216, May 2006.

[30] D. O'Carroll, "Feature-detecting neurons in dragonflies," *Nature*, vol. 362, pp. 541–543, Apr. 1993.

[31] P. T. Gonzalez-Bellido, H. Peng, J. Yang, A. P. Georgopoulos, and R. M. Olberg, "Eight pairs of descending visual neurons in the dragonfly give wing motor centers accurate population vector of prey direction," *Proc. Natl. Acad. Sci. U.S.A. (PNAS)*, vol. 110, pp. 696–701, Jan. 2013.

[32] S. Namiki, M. H. Dickinson, A. M. Wong, W. Korff, and G. M. Card, "The functional organization of descending sensory-motor pathways in drosophila," *Elife*, vol. 7, Jun. 2018, Art. no. e34272.

[33] S. Nicholas, J. Supple, R. Leibbrandt, P. T. Gonzalez-Bellido, and K. Nordström, "Integration of small-and wide-field visual features in target-selective descending neurons of both predatory and nonpredatory dipterans," *J. Neurosci.*, vol. 38, pp. 10725–10733, Dec. 2018.

[34] S. D. Wiederman, P. A. Shoemaker, and D. C. O'Carroll, "A model for the detection of moving targets in visual clutter inspired by insect physiology," *PLoS One*, vol. 3, pp. 1–11, Jul. 2008.

[35] S. D. Wiederman and D. C. OCarroll, "Biologically inspired feature detection using cascaded correlations of off and on channels," *J. Artif. Intell. Soft Comput. Res.*, vol. 3, pp. 5–14, Dec. 2013.

[36] Z. M. Bagheri, B. S. Cazzolato, S. Grainger, D. C. OCarroll, and S. D. Wiederman, "An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments," *J. Neural Eng.*, vol. 14, Jul. 2017, Art. no. 046030.

[37] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. OCarroll, "Performance of an insect-inspired target tracker in natural conditions," *Bioinspiration Biomim.*, vol. 12, Feb. 2017, Art. no. 025006.

[38] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, "Properties of neuronal facilitation that improve target tracking in natural pursuit simulations," *J. Royal Soc. Interface*, vol. 12, Jun. 2015, Art. no. 20150083.

[39] R. D. Santer, F. C. Rind, R. Stafford, and P. J. Simmons, "Role of an identified looming-sensitive neuron in triggering a flying locust's escape," *J. Neurophysiol.*, vol. 95, pp. 3391–3400, Jun. 2006.

[40] M. Y. Peek and G. M. Card, "Comparative approaches to escape," *Curr. Opin. Neurobiol.*, vol. 41, pp. 167–173, Dec. 2016.

[41] M. S. Maisak, J. Haag, G. Ammer, E. Serbe, M. Meier, A. Leonhardt, T. Schilling, A. Bahl, G. M. Rubin, A. Nern, *et al.*, "A directional tuning map of drosophila elementary motion detectors," *Nature*, vol. 500, pp. 212–216, Aug. 2013.

[42] A. Borst and J. Haag, "Neural networks in the cockpit of the fly," *J. Comp. Physiol. A*, vol. 188, pp. 419–437, Jul. 2002.

[43] M. Perry, N. Konstantinides, F. Pinto-Teixeira, and C. Desplan, "Generation and evolution of neural cell types and circuits: insights from the drosophila visual system," *Annual review of genetics*, vol. 51, pp. 501–527, Nov. 2017.

[44] F. C. Rind, "Intracellular characterization of neurons in the locust brain signaling impending collision," *J. Neurophysiol.*, vol. 75, pp. 986–995, Mar. 1996.

[45] P. J. Simmons and F. C. Rind, "Responses to object approach by a wide field visual neurone, the lgmd2 of the locust: characterization and image cues," *J. Comp. Physiol. A*, vol. 180, pp. 203–214, Feb. 1997.

[46] M. O'Shea and J. Williams, "The anatomy and output connection of a locust visual interneurone; the lobular giant movement detector (lgmd) neurone," *J. Comp. Physiol. A*, vol. 91, pp. 257–266, Sep. 1974.

[47] M. Heisenberg and E. Buchner, "The role of retinula cell types in visual behavior ofdrosophila melanogaster," *J. Comp. Physiol.*, vol. 117, pp. 127–162, Jan. 1977.

[48] S. Yamaguchi, R. Wolf, C. Desplan, and M. Heisenberg, "Motion vision is independent of color in drosophila," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 105, pp. 4910–4915, Mar. 2008.

[49] J. Rister and C. Desplan, "The retinal mosaics of opsin expression in invertebrates and vertebrates," *Dev. Neurobiol.*, vol. 71, pp. 1212–1226, May 2011.

[50] S.-Y. Takemura, Z. Lu, and I. A. Meinertzhagen, "Synaptic circuits of the drosophila optic lobe: the input terminals to the medulla," *J. Comp. Neurol.*, vol. 509, pp. 493–513, Jun. 2008.

[51] J. C. Tuthill, A. Nern, S. L. Holtz, G. M. Rubin, and M. B. Reiser, "Contributions of the 12 neuron classes in the fly lamina to motion vision," *Neuron*, vol. 79, pp. 128–140, Jul. 2013.

[52] D. A. Clark and J. B. Demb, "Parallel computations in insect and mammalian visual motion processing," *Curr. Biol.*, vol. 26, pp. R1062–R1072, Oct. 2016.

[53] M. Joesch, B. Schnell, S. V. Raghu, D. F. Reiff, and A. Borst, "On and off pathways in drosophila motion vision," *Nature*, vol. 468, pp. 300–304, Nov. 2010.

[54] D. A. Clark, L. Bursztyn, M. A. Horowitz, M. J. Schnitzer, and T. R. Clandinin, "Defining the computational structure of the motion detector in drosophila," *Neuron*, vol. 70, pp. 1165–1177, Jun. 2011.

[55] M. Silies, D. M. Gohl, Y. E. Fisher, L. Freifeld, D. A. Clark, and T. R. Clandinin, "Modular use of peripheral input channels tunes motion-detecting circuitry," *Neuron*, vol. 79, pp. 111–127, Jul. 2013.

[56] J. Rusanen, A. Vähäkainu, M. Weckström, and K. Arikawa, "Characterization of the first-order visual interneurons in the visual system of the bumblebee (bombus terrestris)," *J. Comp. Physiol. A*, vol. 203, pp. 903–913, Nov. 2017.

[57] J. Li, J. Lindemann, and M. Egelhaaf, "Local motion adaptation enhances the representation of spatial structure at emd arrays," *PLOS Comput. Biol.*, vol. 13, Dec. 2017, Art. no. e1005919.

[58] S.-Y. Takemura, A. Nern, D. B. Chklovskii, L. K. Scheffer, G. M. Rubin, and I. A. Meinertzhagen, "The comprehensive connectome of a neural substrate for onmotion detection in drosophila," *Elife*, vol. 6, Jul. 2017, Art. no. e24394.

[59] M. Joesch, F. Weber, H. Eichner, and A. Borst, "Functional specialization of parallel motion detection circuits in the fly," *J. Neurosci.*, vol. 33, pp. 902–905, Jan. 2013.

[60] A. Arenz, M. S. Drews, F. G. Richter, G. Ammer, and A. Borst, "The temporal tuning of the drosophila motion detectors is determined by the dynamics of their input elements," *Curr. Biol.*, vol. 27, pp. 929–944, Apr. 2017.

[61] E. Serbe, M. Meier, A. Leonhardt, and A. Borst, "Comprehensive characterization of the major presynaptic elements to the drosophila off motion detector," *Neuron*, vol. 89, pp. 829–841, Feb. 2016.

[62] G. Ammer, A. Leonhardt, A. Bahl, B. J. Dickson, and A. Borst, "Functional specialization of neural input elements to the drosophila on motion detector," *Curr. Biol.*, vol. 25, pp. 2247–2253, Aug. 2015.

[63] H. H. Yang, F. St-Pierre, X. Sun, X. Ding, M. Z. Lin, and T. R. Clandinin, "Subcellular imaging of voltage and calcium signals reveals neural processing in vivo," *Cell*, vol. 166, pp. 245–257, Jun. 2016.

[64] K. Shinomiya, G. Huang, Z. Lu, T. Parag, C. S. Xu, R. Aniceto, N. Ansari, N. Cheatham, S. Lauchie, E. Neace, *et al.*, "Comparisons between the on-and off-edge motion pathways in the drosophila brain," *ELife*, vol. 8, Jan. 2019, Art. no. e40025.

[65] J. K. Douglass and N. J. Strausfeld, "Visual motion-detection circuits in flies: parallel direction- and non-direction-sensitive pathways between the medulla and lobula plate," *J. Neurosci.*, vol. 16, pp. 4551–4562, Aug. 1996.

[66] F. C. Rind and P. J. Simmons, "Seeing what is coming: building collision-sensitive neurones," *Trends Neurosci.*, vol. 22, pp. 215–220, May 1999.

[67] F. C. Rind and D. Bramwell, "Neural network based on the input organization of an identified neuron signaling impending collision," *J. Neurophysiol.*, vol. 75, no. 3, pp. 967–985, 1996.

[68] F. C. Rind, S. Wernitznig, P. Pölt, A. Zankel, D. Gütl, J. Sztarker, and G. Leitinger, "Two identified looming detectors in the locust: ubiquitous lateral connections among their inputs contribute to selective responses to looming objects," *Sci. Rep.*, vol. 6, 2016, Art. no. 35525.

[69] R. D. Santer, R. Stafford, and F. C. Rind, "Retinally-generated saccadic suppression of a locust looming-detector neuron: investigations using a robot locust," *J. Royal Soc. Interface*, vol. 1, no. 1, pp. 61–77, 2004.

[70] M. Blanchard, F. C. Rind, and P. F. Verschure, "Collision avoidance using a model of the locust lgmd neuron," *Rob. Auton. Syst.*, vol. 30, pp. 17–38, Jan. 2000.

[71] Q. Fu, C. Hu, J. Peng, and S. Yue, "Shaping the collision selectivity in a looming sensitive neuron model with parallel on and off pathways and spike frequency adaptation," *Neural Netw.*, vol. 106, pp. 127–143, Oct. 2018.

[72] H. Meng, S. Yue, A. Hunter, K. Appiah, M. Hobden, N. Priestley, P. Hobden, and C. Pettit, "A modified neural network model for lobula giant movement detector with additional depth movement feature," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, pp. 2078–2083, Jul. 2009.

[73] A. Silva and C. P. Santos, "Modeling disinhibition within a layered structure of the lgmd neuron," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, pp. 1–7, 2013.

[74] D. Ianchis, V. Tiponut, S. Popescu, and Z. Haraszy, "Improved collision detection system inspired from the neural network of the locust," in *Proc. Int. Symp. Int. Syst. Inf. (SISY)*, pp. 211–215, 2011.

[75] H. Meng, K. Appiah, S. Yue, A. Hunter, M. Hobden, N. Priestley, P. Hobden, and C. Pettit, "A modified model for the lobula giant movement detector and its fpga implementation," *Comput. Vis. Image Underst.*, vol. 114, no. 11, pp. 1238–1247, 2010.

[76] A. Krejan and A. Trost, "Lgmd-based bio-inspired algorithm for detecting risk of collision of a road vehicle," in *Proc. Int. Symp. Image Signal Process. Anal. (ISPA)*, pp. 319–324, IEEE, 2011.

[77] M. Deng, A. Inoue, Y. Shibata, K. Sekiguchi, and N. Ueki, "An obstacle avoidance method for two wheeled mobile robot," in *Proc. Int. Conf. Netw. Sensing Control*, pp. 689–692, IEEE, 2007.

[78] F. Gabbiani, G. Laurent, N. Hatsopoulos, and H. G. Krapp, "The many ways of building collision-sensitive neurons," *Trends Neurosci.*, vol. 22, pp. 437–438, Oct. 1999.

[79] F. Gabbiani, H. G. Krapp, and G. Laurent, "Computation of object approach by a wide-field, motion-sensitive neuron," *J. Neurosci.*, vol. 19, pp. 1122–1141, Feb. 1999.

[80] F. Gabbiani, C. Mo, and G. Laurent, "Invariance of angular threshold computation in a wide-field looming-sensitive neuron," *J. Neurosci.*, vol. 21, pp. 314–329, Jan. 2001.

[81] F. Gabbiani, I. Cohen, and G. Laurent, "Time-dependent activation of feed-forward inhibition in a looming-sensitive neuron," *J. Neurophysiol.*, vol. 94, no. 3, pp. 2150–2161, 2005.

[82] M. S. Keil, "Emergence of multiplication in a biophysical model of a wide-field visual neuron for computing object approaches: Dynamics, peaks, & fits," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, pp. 469–477, Curran Associates, Inc., 2011.

[83] M. S. Keil, "Dendritic pooling of noisy threshold processes can explain many properties of a collision-sensitive visual neuron," *PLOS Comput. Biol.*, vol. 11, no. 10, 2015, Art. no. e1004479.

[84] S. B. i Badia and P. F. Verschure, "A collision avoidance model based on the lobula giant movement detector (lgmd) neuron of the locust," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, vol. 3, pp. 1757–1761, IEEE, 2004.

[85] S. B. i Badia, U. Bernardet, and P. F. Verschure, "Non-linear neuronal responses as an emergent property of afferent networks: A case study of the locust lobula giant movement detector," *PLoS computational biology*, vol. 6, no. 3, 2010, Art. no. e1000701.

[86] F. Gabbiani, H. G. Krapp, N. Hatsopoulos, C.-H. Mo, C. Koch, and G. Laurent, "Multiplication and stimulus invariance in a looming-sensitive neuron," *J. Physiol. Paris*, vol. 98, no. 1-3, pp. 19–34, 2004.

[87] R. Stafford, R. D. Santer, and F. C. Rind, "A bio-inspired visual collision detection mechanism for cars: combining insect inspired neurons to create a robust system," *BioSystems*, vol. 87, no. 2-3, pp. 164–171, 2007.

[88] K. Hausen, "The lobula-complex of the fly: structure, function and significance in visual behaviour," in *Photoreception and vision in invertebrates*, pp. 523–559, Springer, 1984.

[89] J. A. Strother, S.-T. Wu, E. M. Rogers, J. L. Eliason, A. M. Wong, A. Nern, and M. B. Reiser, "Behavioral state modulates the on visual motion pathway of drosophila," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 115, pp. E102–E111, Jan. 2018.

[90] H. G. Krapp, B. Hengstenberg, and R. Hengstenberg, "Dendritic structure and receptive-field organization of optic flow processing interneurons in the fly," *J. Neurophysiol.*, vol. 79, pp. 1902–1917, Apr. 1998.

[91] F. Iida and D. Lambrinos, "Navigation in an autonomous flying robot by using a biologically inspired visual odometer," in *Sensor Fusion and Decentralized Control in Robotic Systems III*, vol. 4196, pp. 86–98, International Society for Optics and Photonics, 2000.

[92] B. Hassenstein and W. Reichardt, "Systemtheoretische analyse der zeit-, reihenfolgen-und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers chlorophanus," *Zeitschrift für Naturforschung B*, vol. 11, pp. 513–524, Oct. 1956.

[93] N. Franceschini, A. Riehle, and A. Le Nestour, "Directionally selective motion detection by insect neurons," in *Facets of vision*, pp. 360–390, Berlin, Heidelberg: Springer, 1989.

[94] M. Egelhaaf and A. Borst, "Are there separate on and off channels in fly motion vision?," *Vis. Neurosci.*, vol. 8, no. 2, pp. 151–164, 1992.

[95] J. C. Tuthill, M. E. Chiappe, and M. B. Reiser, "Neural correlates of illusory motion perception in drosophila," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, pp. 9685–9690, Jun. 2011.

[96] A. J. Cope, C. Sabo, K. Gurney, E. Vasilaki, and J. A. Marshall, "A model for an angular velocity-tuned motion detector accounting for deviations in the corridor-centering response of the bee," *PLOS Comput. Biol.*, vol. 12, May 2016, Art. no. e1004887.

[97] F. Iida, "Biologically inspired visual odometer for navigation of a flying robot," *Rob. Auton. Syst.*, vol. 44, pp. 201–208, Sep. 2003.

[98] S. Rajesh, D. OCarroll, and D. Abbott, "Man-made velocity estimators based on insect vision," *Smart Mater. Struct.*, vol. 14, p. 413, Mar. 2005.

[99] O. J. Bertrand, J. P. Lindemann, and M. Egelhaaf, "A bio-inspired collision avoidance model based on spatial information derived from motion detectors leads to common routes," *PLOS Comput. Biol.*, vol. 11, Nov. 2015, Art. no. e1004339.

[100] Z. Zhang, S. Yue, and G. Zhang, "Fly visual system inspired artificial neural network for collision detection," *Neurocomputing*, vol. 153, pp. 221–234, Apr. 2015.

[101] J. M. Missler and F. A. Kamangar, "A neural network for pursuit tracking inspired by the fly visual system," *Neural Netw.*, vol. 8, no. 3, pp. 463–480, 1995.

[102] V. Pant and C. M. Higgins, "Tracking improves performance of biological collision avoidance models," *Biol. Cybern.*, vol. 106, pp. 307–322, Jul. 2012.

[103] P. Sterling and B. G. Wickelgren, "Visual receptive fields in the superior colliculus of the cat.," *J. Neurophysiol.*, vol. 32, no. 1, pp. 1–15, 1969.

[104] G. Rizzolatti and R. Camarda, "Influence of the presentation of remote visual stimuli on visual responses of cat area 17 and lateral suprasylvian area," *Exp. Brain Res.*, vol. 29, pp. 107–122, Aug. 1977.

[105] G. Mandl, "Responses of visual cells in cat superior colliculus to relative pattern movement," *Vision Res.*, vol. 25, no. 2, pp. 267–281, 1985.

[106] R. Davidson and D. Bender, "Selectivity for relative motion in the monkey superior colliculus," *J. Neurophysiol.*, vol. 65, pp. 1115–1133, May 1991.

[107] K. Tanaka, K. Hikosaka, H.-a. Saito, M. Yukie, Y. Fukada, and E. Iwai, "Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey," *J. Neurosci.*, vol. 6, pp. 134–144, Jan. 1986.

[108] J. Allman, F. Miezin, and E. McGuinness, "Direction-and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (mt)," *Perception*, vol. 14, no. 2, pp. 105–126, 1985.

[109] B. Frost and K. Nakayama, "Single visual neurons code opposing motion independent of direction," *Science*, vol. 220, pp. 744–745, May 1983.

[110] B. Frost, P. Cavanagh, and B. Morgan, "Deep tectal cells in pigeons respond to kinematograms," *J. Comp. Physiol. A*, vol. 162, pp. 639–647, Sep. 1988.

[111] B. Frost, P. Scilley, and S. Wong, "Moving background patterns reveal double-opponency of directionally specific pigeon tectal neurons," *Exp. Brain Res.*, vol. 43, pp. 173–185, Jul. 1981.

[112] H.-J. Tsai, "Responses of toad's tectal neurons to in-phase and anti-phase movements of object and textured background," *J. Comp. Physiol. A*, vol. 167, pp. 857–863, Dec. 1990.

[113] C. F. Rowell and M. O'Shea, "Neuronal basis of a sensory analyser, the acridid movement detector system. iii. control of response amplitude by tonic lateral inhibition," *J. Exp. Biol.*, vol. 65, no. 3, pp. 617–625, 1976.

[114] C. F. Rowell, M. O'Shea, and J. Williams, "The neuronal basis of a sensory analyser, the acridid movement detector system. iv. the preference for small field stimuli," *J. Exp. Biol.*, vol. 68, no. 1, pp. 157–185, 1977.

[115] G. A. Horridge, "Compound eye and vision of insects," in *International Congress of Entomology, Canberra, 1972.*, pp. 437–466, Clarendon Press, 1975.

[116] T. Collett, "Visual neurones for tracking moving targets," *Nature*, vol. 232, pp. 127–130, Jul. 1971.

[117] T. Collett, "Visual neurones in the anterior optic tract of the privet hawk moth," *J. Comp. Physiol.*, vol. 78, pp. 396–433, Dec. 1972.

[118] R. M. Olberg, "Object-and self-movement detectors in the ventral nerve cord of the dragonfly," *J. Comp. Physiol.*, vol. 141, pp. 327–334, Sep. 1981.

[119] R. M. Olberg, "Identified target-selective visual interneurons descending from the dragonfly brain," *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.*, vol. 159, pp. 827–840, Nov. 1986.

[120] V. Gauck and A. Borst, "Spatial response properties of contralateral inhibited lobula plate tangential cells in the fly visual system," *J. Comp. Neurol.*, vol. 406, pp. 51–71, Feb. 1999.

[121] M. Egelhaaf, "On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly. 3: Possible input circuitries and behavioural significance of the fd-cells," *Biol. Cybern.*, vol. 52, no. 4, pp. 267–280, 1985.

[122] B. Kimmerle and M. Egelhaaf, "Detection of object motion by a fly neuron during simulated flight," *J. Comp. Physiol. A*, vol. 186, pp. 21–31, Jan. 2000.

[123] B. Kimmerle, A.-K. Warzecha, and M. Egelhaaf, "Object detection in the fly during simulated translatory flight," *J. Comp. Physiol. A*, vol. 181, pp. 247–255, Aug. 1997.

[124] J. L. Fox, J. W. Aptekar, N. M. Zolotova, P. A. Shoemaker, and M. A. Frye, "Figure–ground discrimination behavior in drosophila. i. spatial organization of wing-steering responses," *J. Exp. Biol.*, vol. 217, pp. 558–569, Feb. 2014.

[125] M. Egelhaaf, "On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly. 2: Figure-detection cells a new class of visual interneurons," *Biol. Cybern.*, vol. 52, no. 2, pp. 123–140, 1985.

[126] M. Egelhaaf, "On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly. 3: Possible input circuitries and behavioural significance of the fd-cells," *Biol. Cybern.*, vol. 52, no. 4, pp. 267–280, 1985.

[127] B. P. Ölveczky, S. A. Baccus, and M. Meister, "Segregation of object and background motion in the retina," *Nature*, vol. 423, p. 401, May 2003.

[128] M. Egelhaaf, K. Hausen, W. Reichardt, and C. Wehrhahn, "Visual course control in flies relies on neuronal computation of object and background motion," *Trends Neurosci.*, vol. 11, no. 8, pp. 351–358, 1988.

[129] B. J. Duistermars, M. B. Reiser, Y. Zhu, and M. A. Frye, "Dynamic properties of large-field and small-field optomotor flight responses in drosophila," *J. Comp. Physiol. A*, vol. 193, pp. 787–799, Jul. 2007.

[130] J.-M. Mongeau and M. A. Frye, "Drosophila spatiotemporally integrates visual signals to control saccades," *Curr. Biol.*, vol. 27, pp. 2901–2914, Oct. 2017.

[131] M.-J. Escobar, D. Pezo, and P. Orio, "Mathematical analysis and modeling of motion direction selectivity in the retina," *J. Physiol. Paris*, vol. 107, pp. 349–359, Nov. 2013.

[132] H. Wang, J. Peng, and S. Yue, "A directionally selective small target motion detecting visual neural network in cluttered backgrounds," *IEEE Trans. Cybern.*, in press, doi: 10.1109/TCYB.2018.2869384.

[133] H. Wang, J. Peng, X. Zheng, and S. Yue, "A robust visual system for small target motion detection against cluttered moving backgrounds," *IEEE Trans. Neural Netw. Learn. Syst.*, in press, doi: 10.1109/TNNLS.2019.2910418.

[134] H. Wang, J. Peng, and S. Yue, "A feedback neural network for small target motion detection in cluttered backgrounds," in *Proc. Int. Conf. Artif. Neural Netw. (ICANN)*, pp. 728–737, Springer, 2018.

[135] A.-K. Warzecha, M. Egelhaaf, and A. Borst, "Neural circuit tuning fly visual interneurons to motion of small objects. i. dissection of the circuit by pharmacological and photoinactivation techniques," *J. Neurophysiol.*, vol. 69, no. 2, pp. 329–339, 1993.

[136] H. G. Krapp, R. Hengstenberg, and M. Egelhaaf, "Binocular contributions to optic flow processing in the fly visual system," *J. Neurophysiol.*, vol. 85, no. 2, pp. 724–734, 2001.

[137] C. Spalthoff, M. Egelhaaf, P. Tinnefeld, and R. Kurtz, "Localized direction selective responses in the dendrites of visual interneurons of the fly," *BMC Biol.*, vol. 8, no. 1, p. 36, 2010.

[138] H. Eckert and D. R. Dvorak, "The centrifugal horizontal cells in the lobula plate of the blowfly, phaenicia sericata," *J. Insect Physiol.*, vol. 29, no. 7, pp. 547–560, 1983.

[139] W. Reichardt, T. Poggio, and K. Hausen, "Figure-ground discrimination by relative movement in the visual system of the fly," *Biol. Cybern.*, vol. 46, pp. 1–30, Jan. 1983.

[140] W. Reichardt, M. Egelhaaf, and A.-k. Guo, "Processing of figure and background motion in the visual system of the fly," *Biol. Cybern.*, vol. 61, pp. 327–345, Sep. 1989.

[141] C. M. Higgins and V. Pant, "An elaborated model of fly small-target tracking," *Biol. Cybern.*, vol. 91, pp. 417–428, Dec. 2004.

[142] D. Fortun, P. Bouthemy, and C. Kervrann, "Optical flow modeling and computation: a survey," *Comput. Vis. Image Underst.*, vol. 134, pp. 1–21, May 2015.

[143] H. Yong, D. Meng, W. Zuo, and L. Zhang, "Robust online matrix factorization for dynamic background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, pp. 1726–1740, Jul. 2017.

[144] Z. Li, G. Zhao, S. Li, H. Sun, R. Tao, X. Huang, and Y. J. Guo, "Rotation feature extraction for moving targets based on temporal differencing and image edge detection.," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, pp. 1512–1516, Oct. 2016.

[145] Y. Ren, C.-S. Chua, and Y.-K. Ho, "Motion detection with nonstationary background," *Mach. Vis. Appl.*, vol. 13, pp. 332–343, Mar. 2003.

[146] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, pp. 185–203, Aug. 1981.

[147] J. J. Gibson, *The perception of the visual world.* Oxford, England: Houghton Mifflin, 1950.

[148] E. Mémin and P. Pérez, "Hierarchical estimation and segmentation of dense motion fields," *Int. J. Comput. Vision*, vol. 46, pp. 129–155, Feb 2002.

[149] M. J. Black, Y. Yacoob, A. D. Jepson, and D. J. Fleet, "Learning parameterized models of image motion," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 561–567, IEEE, 1997.

[150] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," *IEEE Trans Med Imaging*, vol. 18, pp. 712–721, Aug. 1999.

[151] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proc. Conf. ECCV*, pp. 237–252, Springer, 1992.

[152] X. Shen and Y. Wu, "Sparsity model for robust optical flow estimation at motion discontinuities," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2456–2463, IEEE, 2010.

[153] H. Zimmer, A. Bruhn, and J. Weickert, "Optic flow in harmony," *Int. J. Comput. Vision*, vol. 93, pp. 368–388, Jul. 2011.

[154] J. Weickert and C. Schnörr, "A theoretical framework for convex regularizers in pde-based computation of image motion," *Int. J. Comput. Vision*, vol. 45, pp. 245–264, Dec. 2001.

[155] F. Heitz and P. Bouthemy, "Multimodal estimation of discontinuous optical flow using markov random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, pp. 1217–1232, Dec. 1993.

[156] M. Piccardi, "Background subtraction techniques: a review," in *IEEE Int. Conf. Syst. Man Cybern.*, vol. 4, pp. 3099–3104, IEEE, 2004.

[157] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, pp. 246–252, IEEE, 1999.

[158] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, "On the applications of robust pca in image and video processing," *Proc. IEEE*, vol. 106, pp. 1427–1457, Aug. 2018.

[159] E. Hayman and J.-O. Eklundh, "Statistical background subtraction for a mobile observer," in *Proc. IEEE Int. Conf. Comp. Vis. (ICCV)*, p. 67, IEEE, 2003.

[160] Z. Zivkovic and F. Van Der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.*, vol. 27, pp. 773–780, May 2006.

[161] K. Moo Yi, K. Yun, S. Wan Kim, H. Jin Chang, and J. Young Choi, "Detection of moving objects with non-stationary cameras in 5.8 ms: Bringing motion detection to your mobile device," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, pp. 27–34, 2013.

[162] Y.-T. Chen, C.-S. Chen, C.-R. Huang, and Y.-P. Hung, "Efficient hierarchical method for background subtraction," *Pattern Recognit.*, vol. 40, pp. 2706–2715, Oct. 2007.

[163] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Comput. Sci. Rev*, vol. 23, pp. 1–71, Feb. 2017.

[164] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM (JACM)*, vol. 58, p. 11, May 2011.

[165] Z. Gao, L.-F. Cheong, and Y.-X. Wang, "Block-sparse rpca for salient motion detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, pp. 1975–1987, Apr. 2014.

[166] B. Xin, Y. Tian, Y. Wang, and W. Gao, "Background subtraction via generalized fused lasso foreground modeling," in *Proc. IEEE CVPR*, pp. 4676–4684, Jun. 2015.

[167] S. Javed, A. Mahmood, S. Al-Maadeed, T. Bouwmans, and S. K. Jung, "Moving object detection in complex scene using spatiotemporal structured-sparse rpca," *IEEE Trans. Image Process.*, vol. 28, pp. 1007–1022, Feb. 2019.

[168] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, pp. 597–610, Mar. 2013.

[169] C. Chen, S. Li, H. Qin, and A. Hao, "Robust salient motion detection in non-stationary videos via novel integrated strategies of spatio-temporal coherency clues and low-rank analysis," *Pattern Recognit.*, vol. 52, pp. 410–432, Apr. 2016.

[170] S. E. Ebadi, V. G. Ones, and E. Izquierdo, "Efficient background subtraction with low-rank and sparse matrix decomposition," in *IEEE Int. Conf. Image Proc. (ICIP)*, pp. 4863–4867, IEEE, 2015.

[171] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE Trans. Image Process.*, vol. 14, pp. 294–307, Mar. 2005.

[172] I. Saleemi and M. Shah, "Multiframe many–many point correspondence for vehicle tracking in high density wide area aerial videos," *Int. J. Comput. Vis.*, vol. 104, pp. 198–219, Sep. 2013.

[173] D. Zhou and H. Zhang, "Modified gmm background modeling and optical flow for detection of moving objects," in *IEEE Int. Conf. Syst. Man Cybern.*, vol. 3, pp. 2224–2229, IEEE, 2005.

[174] J. Xiao, H. Cheng, H. Sawhney, and F. Han, "Vehicle detection and tracking in wide field-of-view aerial video," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 679–684, IEEE, 2010.

[175] G. Saur, W. Krüger, and A. Schumann, "Extended image differencing for change detection in uav video mosaics," in *Proceedings of SPIE*, vol. 9026, 2014.

[176] G. Indiveri and R. Douglas, "Neuromorphic vision sensors," *Science*, vol. 288, pp. 1189–1190, May 2000.

[177] F. C. Rind, R. D. Santer, J. M. Blanchard, and P. F. Verschure, "Locusts looming detectors for robot sensors," in *Sensors and sensing in biology and engineering*, pp. 237–250, Springer, 2003.

[178] Y. Gu, C. Wang, B. Liu, and Y. Zhang, "A kernel-based nonparametric regression method for clutter removal in infrared small-target detection applications," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, pp. 469–473, Feb. 2010.

[179] J.-F. Rivest and R. Fortin, "Detection of dim targets in digital infrared imagery by morphological image processing," *Opt. Eng.*, vol. 35, pp. 1886–1894, Jul. 1996.

[180] K. L. Anderson and R. A. Iltis, "A tracking algorithm for infrared images based on reduced sufficient statistics," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 464–472, Apr. 1997.

[181] Y. Xiong, J.-X. Peng, M.-Y. Ding, and D.-H. Xue, "An extended track-before-detect algorithm for infrared target detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 1087–1092, Jul. 1997.

[182] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Infrared small-target detection using multiscale gray difference weighted image entropy," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, pp. 60–72, Apr. 2016.

[183] S. D. Deshpande, M. H. Er, R. Venkateswarlu, and P. Chan, "Max-mean and max-median filters for detection of small targets," in *Proc. SPIE*, vol. 3809, pp. 74–84, Oct. 1999.

[184] T.-W. Bae, "Small target detection using bilateral filter and temporal cross product in infrared images," *Infr. Phys. Technol.*, vol. 54, pp. 403–411, Sep. 2011.

[185] K. Shang, X. Sun, J. Tian, Y. Li, and J. Ma, "Infrared small target detection via line-based reconstruction and entropy-induced suppression," *Infrared Phys. Technol.*, vol. 76, pp. 75–81, May 2016.

[186] M. M. Hadhoud and D. W. Thomas, "The two-dimensional adaptive lms (tdlms) algorithm," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 485–494, May 1988.

[187] M. R. Azimi-Sadjadi and H. Pan, "Two-dimensional block diagonal lms adaptive filtering," *IEEE Trans. Signal Process.*, vol. 42, pp. 2420–2429, Sep. 1994.

[188] P. A. Ffrench, J. Zeidler, and W. H. Ku, "Enhanced detectability of small objects in correlated clutter using an improved 2-d adaptive lattice algorithm," *IEEE Trans. Image Process.*, vol. 6, pp. 383–397, Mar. 1997.

[189] Y. Cao, R. Liu, and J. Yang, "Small target detection using two-dimensional least mean square (tdlms) filter based on neighborhood analysis," *Int. J. Infr. Millimeter Waves*, vol. 29, no. 2, pp. 188–200, 2008.

[190] V. T. Tom, T. Peli, M. Leung, and J. E. Bondaryk, "Morphology-based algorithm for point target detection in infrared backgrounds," in *Proc. SPIE*, vol. 1954, pp. 2–12, Oct. 1993.

[191] A. Toet and T. Wu, "Small maritime target detection through false color fusion," in *Proc. SPIE*, vol. 6945, pp. 69450V–69453V, Apr. 2008.

[192] X. Bai, F. Zhou, and B. Xue, "Infrared dim small target enhancement using toggle contrast operator," *Infrared Phys. Technol.*, vol. 55, pp. 177–182, Mar. 2012.

[193] X. Bai, F. Zhou, and B. Xue, "Fusion of infrared and visual images through region extraction by using multi scale center-surround top-hat transform," *Opt. Express*, vol. 19, pp. 8444–8457, Apr. 2011.

[194] X. Bai and F. Zhou, "Hit-or-miss transform based infrared dim small target enhancement," *Opt. Laser Technol.*, vol. 43, pp. 1084–1090, Oct. 2011.

[195] W. Meng, T. Jin, and X. Zhao, "Adaptive method of dim small object detection with heavy clutter," *Appl. Opt.*, vol. 52, pp. D64–D74, Apr. 2013.

[196] B. Xiangzhi, Z. Fugen, X. Yongchun, and J. Ting, "Modified top-hat transformation based on contour structuring element to detect infrared small target," in *Proc. 3rd IEEE Conf. ICIEA*, pp. 575–579, IEEE, Jun. 2008.

[197] X. Bai and F. Zhou, "Analysis of different modified top-hat transformations based on structuring element construction," *Signal Process.*, vol. 90, pp. 2999–3003, Nov. 2010.

[198] G.-D. Wang, C.-Y. Chen, and X.-B. Shen, "Facet-based infrared small target detection method," *Electron. Lett.*, vol. 41, pp. 1244–1246, Nov. 2005.

[199] P. Wang, J. Tian, and C. Q. Gao, "Infrared small target detection using directional highpass filters based on ls-svm," *Electron. Lett.*, vol. 45, pp. 156–158, Jan. 2009.

[200] C. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, pp. 574–581, Jan. 2014.

[201] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, and Y. Fang, "A robust infrared small target detection algorithm based on human visual system," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2168–2172, 2014.

[202] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Small infrared target detection based on weighted local difference measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, pp. 4204–4214, Jul. 2016.

[203] Y. Wei, X. You, and H. Li, "Multiscale patch-based contrast measure for small infrared target detection," *Pattern Recognit.*, vol. 58, pp. 216–226, Oct. 2016.

[204] S. Kim, "Min-local-log filter for detecting small targets in cluttered background," *Electron. Lett.*, vol. 47, pp. 105–106, Jan. 2011.

[205] Y. He, M. Li, J. Zhang, and Q. An, "Small infrared target detection based on low-rank and sparse representation," *Infrared Phys. Technol.*, vol. 68, pp. 98–109, Jan. 2015.

[206] X. Wang, Z. Peng, D. Kong, and Y. He, "Infrared dim and small target detection based on stable multisubspace learning in heterogeneous scene," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, pp. 5481–5493, Oct. 2017.

[207] I. S. Reed, R. M. Gagliardi, and L. B. Stotts, "Optical moving target detection with 3-d matched filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, pp. 327–336, Jul. 1988.

[208] B. Porat and B. Friedlander, "A frequency domain algorithm for multiframe detection and estimation of dim targets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, pp. 398–401, Apr. 1990.

[209] M. Li, T. Zhang, W. Yang, and X. Sun, "Moving weak point target detection and estimation with three-dimensional double directional filter in ir cluttered background," *Opt. Eng.*, vol. 44, no. 10, p. 107007, 2005.

[210] X. Liu and Z. Zuo, "A dim small infrared moving target detection algorithm based on improved three-dimensional directional filtering," in *Chinese Conference on Image and Graphics Technologies*, vol. 363, pp. 102–108, Springer-Verlag, 2013.

[211] A. P. Tzannes and D. H. Brooks, "Detecting small moving objects using temporal hypothesis testing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, pp. 570–586, Aug. 2002.

[212] E. T. Lim, L. Shue, and R. Venkateswarlu, "Adaptive mean and variance filter for detection of dim point-like targets," *Proc. SPIE*, vol. 4728, pp. 492–503, Aug. 2002.

[213] D. Liu, J. Zhang, and W. Dong, "Temporal profile based small moving target detection algorithm in infrared image sequences," *Int. J. Infr. Millimeter Waves*, vol. 28, pp. 373–381, May 2007.

[214] Z. Wang, J. Tian, J. Liu, and S. Zheng, "Small infrared target fusion detection based on support vector machines in the wavelet domain," *Opt. Eng.*, vol. 45, p. 076401, Jul. 2006.

[215] B. Zhang, T. Zhang, Z. Cao, and K. Zhang, "Fast new small-target detection algorithm based on a modified partial differential equation in infrared clutter," *Opt. Eng.*, vol. 46, p. 106401, Oct. 2007.

[216] E. J. Warrant, "Matched filtering and the ecology of vision in insects," in *The Ecology of Animal Senses*, pp. 143–167, Cham,Switzerland: Springer, 2016.

[217] L. Freifeld, D. A. Clark, M. J. Schnitzer, M. A. Horowitz, and T. R. Clandinin, "Gabaergic lateral interactions tune the early stages of visual processing in drosophila," *Neuron*, vol. 78, pp. 1075–1089, Jun. 2013.

[218] B. De Vries and J. C. Príncipe, "A theory for neural networks with time delays," in *Proc. NIPS*, pp. 162–168, 1990.

[219] J. C. Principe, B. De Vries, and P. G. De Oliveira, "The gamma-filter-a new class of adaptive iir filters with restricted feedback," *EEE Trans. Signal Process.*, vol. 41, pp. 649–656, Feb. 1993.

[220] B. De Vries and J. C. Principe, "The gamma modela new neural model for temporal processing," *Neural Netw.*, vol. 5, no. 4, pp. 565–576, 1992.

[221] J. C. Tuthill and B. G. Borghuis, "Four to foxtrot: how visual motion is computed in the fly brain," *Neuron*, vol. 89, pp. 677–680, Feb. 2016.

[222] Y.-J. Lee, H. O. Jönsson, and K. Nordström, "Spatio-temporal dynamics of impulse responses to figure motion in optic flow neurons," *PLoS One*, vol. 10, pp. 1–16, May 2015.

[223] A. Borst and M. Egelhaaf, "Direction selectivity of blowfly motion-sensitive neurons is computed in a two-stage process," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 87, pp. 9363–9367, Dec. 1990.

[224] A. Borst, M. Egelhaaf, and J. Haag, "Mechanisms of dendritic integration underlying gain control in fly motion-sensitive interneurons," *J. Comput. Neurosci.*, vol. 2, pp. 5–18, Mar. 1995.

[225] A. D. Straw, "Vision egg: an open-source library for realtime visual stimulus generation," *Front. Neuroinf.*, vol. 2, p. 4, Nov. 2008.

[226] N. Vogt and C. Desplan, "The first steps in drosophila motion detection," *Neuron*, vol. 56, pp. 5–7, Oct. 2007.

[227] L. Zheng, G. G. de Polavieja, V. Wolfram, M. H. Asyali, R. C. Hardie, and M. Juusola, "Feedback network controls photoreceptor output at the layer of first visual synapses in drosophila," *J. Gen. Physiol.*, vol. 127, pp. 495–510, Apr. 2006.

[228] T. Hwu, J. Isbell, N. Oros, and J. Krichmar, "A self-driving robot using deep convolutional neural networks on neuromorphic hardware," in *Proc. Int. Joint Conf. Neural Netw.*, pp. 635–641, 2017.

[229] M. Beyeler, N. Oros, N. Dutt, and J. L. Krichmar, "A gpu-accelerated cortical neural network model for visually guided robot navigation," *Neural Netw.*, vol. 72, pp. 75–87, Dec. 2015.

[230] N. A. Browning, S. Grossberg, and E. Mingolla, "Cortical dynamics of navigation and steering in natural scenes: Motion-based object segmentation, heading, and obstacle avoidance," *Neural Netw.*, vol. 22, pp. 1383–1398, Dec. 2009.

[231] B. Hu, S. Yue, and Z. Zhang, "A rotational motion perception neural network based on asymmetric spatiotemporal visual information processing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, pp. 2803–2821, Nov 2016.

[232] A. L. Stckl, W. A. Ribi, and E. J. Warrant, "Adaptations for nocturnal and diurnal vision in the hawkmoth lamina," *J. Comp. Neurol.*, vol. 524, p. 160, Jul. 2016.

[233] Z. Riveraalvidrez, I. Lin, and C. M. Higgins, "A neuronally based model of contrast gain adaptation in fly motion vision," *Visual Neurosci.*, vol. 28, p. 419, Aug. 2011.

[234] N. Lessios, R. L. Rutowski, J. H. Cohen, M. E. Sayre, and N. J. Strausfeld, "Multiple spectral channels in branchiopods. i. vision in dim light and neural correlates," *J. Exp. Biol.*, pp. jeb–165860, Apr. 2018.

[235] P. Ardin, F. Peng, M. Mangan, K. Lagogiannis, and B. Webb, "Using an insect mushroom body circuit to encode route memory in complex natural environments," *PLoS Comput. Biol.*, vol. 12, pp. 1–12, Feb. 2016.

[236] B. Webb and A. Wystrach, "Neural mechanisms of insect navigation," *Curr. Opin. Insect Sci.*, vol. 15, pp. 27–39, Jun. 2016.

[237] S.-y. Takemura, A. Bharioke, Z. Lu, A. Nern, S. Vitaladevuni, P. K. Rivlin, W. T. Katz, D. J. Olbris, S. M. Plaza, P. Winston, *et al.*, "A visual motion detection circuit suggested by drosophila connectomics," *Nature*, vol. 500, p. 175, Aug. 2013.

[238] R. Behnia and C. Desplan, "Visual circuits in flies: beginning to see the whole picture," *Curr. Opin. Neurobiol.*, vol. 34, pp. 125–132, Oct. 2015.

[239] M. F. Keleş and M. A. Frye, "Object-detecting neurons in drosophila," *Curr. Biol.*, vol. 27, pp. 680–687, Mar. 2017.

[240] S. Yamaguchi and M. Heisenberg, "Photoreceptors and neural circuitry underlying phototaxis in insects," *Fly*, vol. 5, pp. 333–336, Dec. 2011.

[241] S. M. Rogers and S. R. Ott, "Differential activation of serotonergic neurons during short-and long-term gregarization of desert locusts," *Proc. Royal Soc. B*, vol. 282, p. 20142062, Feb. 2015.

[242] W. T. Freeman, E. H. Adelson, *et al.*, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, pp. 891–906, Sep. 1991.

[243] W.-C. Zhang and P.-L. Shui, "Contour-based corner detection via angle difference of principal directions of anisotropic gaussian directional derivatives," *Pattern Recognit.*, vol. 48, pp. 2785–2797, Sep. 2015.

[244] H. Kafaligonul, B. G. Breitmeyer, and H. Öğmen, "Feedforward and feedback processes in vision," *Front. Psychol.*, vol. 6, p. 279, Mar. 2015.

[245] V. A. Lamme, H. Super, and H. Spekreijse, "Feedforward, horizontal, and feedback processing in the visual cortex," *Curr. Opin. Neurobiol.*, vol. 8, pp. 529–535, Aug. 1998.

[246] Y. Mohsenzadeh, S. Qin, R. M. Cichy, and D. Pantazis, "Ultra-rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway," *Elife*, vol. 7, p. e36329, Jun. 2018.

[247] J. Hupé, A. James, B. Payne, S. Lomber, P. Girard, and J. Bullier, "Cortical feedback improves discrimination between figure and background by v1, v2 and v3 neurons," *Nature*, vol. 394, p. 784, Aug. 1998.

[248] A. Pavan, L. M. Bimson, M. G. Gall, F. Ghin, and G. Mather, "The interaction between orientation and motion signals in moving oriented glass patterns," *Vis. Neurosci.*, vol. 34, Jul. 2017.

[249] P. C. Klink, B. Dagnino, M.-A. Gariel-Mathis, and P. R. Roelfsema, "Distinct feedforward and feedback effects of microstimulation in visual cortex reveal neural mechanisms of texture segregation," *Neuron*, vol. 95, pp. 209–220, Jul. 2017.

[250] L. Nurminen, S. Merlin, M. Bijanzadeh, F. Federer, and A. Angelucci, "Top-down feedback controls spatial summation and response amplitude in primate visual cortex," *Nat. Commun.*, vol. 9, p. 2281, Jun. 2018.

[251] B. Gulyas, G. Orban, J. Duysens, and H. Maes, "The suppressive influence of moving textured backgrounds on responses of cat striate neurons to moving bars," *J. Neurophysiol.*, vol. 57, pp. 1767–1791, Jun. 1987.

[252] C. Trischler, R. Kern, and M. Egelhaaf, "Chasing behaviour and optomotor following in free-flying male blowflies: flight performance and interactions of the underlying control systems," *Front. Behav. Neurosci.*, vol. 4, p. 20, May 2010.

[253] M. Mitchell, *An introduction to genetic algorithms*. MIT press, 1998.

[254] M. Gen and L. Lin, "Genetic algorithms," in *Wiley Encyclopedia of Computer Science and Engineering*, pp. 1–15, Wiley Online Library, Jun. 2007.

[255] O. Dyakova, M. M. Müller, M. Egelhaaf, and K. Nordström, "Image statistics of the environment surrounding freely behaving hoverflies," *J. Comp. Physiol. A*, pp. 1–13, Apr. 2019.